Article

Improving Signal-Strength Aggregation for Mobile Crowdsourcing Scenarios

Diego Madariaga [1,2,*] , Javier Madariaga [1] , Javier Bustos-Jiménez [1] and Benjamin Bustos [2]

1 NIC Chile Research Labs, University of Chile, Santiago 8320000, Chile; javier@niclabs.cl (J.M.); jbustos@niclabs.cl (J.B.-J.)

2 Department of Computer Science, Millennium Institute Foundational Research on Data, University of Chile, Santiago 8320000, Chile; bebustos@dcc.uchile.cl

* Correspondence: diego@niclabs.cl

Abstract: Due to its huge impact on the overall quality of service (QoS) of wireless networks, both academic and industrial research have actively focused on analyzing the received signal strength in areas of particular interest. In this paper, we propose the improvement of signal-strength aggregation with a special focus on Mobile Crowdsourcing scenarios by avoiding common issues related to the mishandling of log-scaled signal values, and by the proposal of a novel aggregation method based on interpolation. Our paper presents two clear contributions. First, we discuss the misuse of log-scaled signal-strength values, which is a persistent problem within the mobile computing community. We present the physical and mathematical formalities on how signal-strength values must be handled in a scientific environment. Second, we present a solution to the difficulties of aggregating signal strength in Mobile Crowdsourcing scenarios, as a low number of measurements and nonuniformity in spatial distribution. Our proposed method obtained consistently lower Root Mean Squared Error (RMSE) values than other commonly used methods at estimating the expected value of signal strength over an area. Both contributions of this paper are important for several recent pieces of research that characterize signal strength for an area of interest.

Keywords: signal strength; mobile crowdsourcing; network measurements

1. Introduction

During the last decade, many research studies have made use of Mobile Crowdsourcing methods to analyze the performance and quality of service (QoS) in mobile environments. These studies usually obtain different QoS indicators together with some environmental data such as timestamps, location coordinates and cell identifiers, to describe wireless network behavior for a given geographical area. Among all the collected network information, the received signal-strength indicator is included in most Mobile Crowdsourcing analyses. This recurrent consideration of signal strength is in part because it is very easy to obtain from end-user mobile devices [1], but mostly because of its influence on the overall QoS in wireless networks, which is reflected on the impact produced by signal-strength variations in network performance measurements [1–4]. Moreover, analyses over signal-strength data are not only interesting for academic research, but also for mobile analytics companies as *OpenSignal* and *Tutela*, and for mobile network operators for radio network planning [5,6] and for performing coverage analysis in cellular networks [7].

A common methodology to summarize the received signal strength inside a specific area is to aggregate all the individual measurements into one representative value that characterizes the signal strength inside the location area [1,8–18].

The first question about getting a representative signal-strength value from the aggregation of several individual measurements is to identify what real value we actually want to represent and estimate. In this paper, we consider the formal definition of the expected value of signal strength as the target value to be estimated. The expected value is

a measure of central tendency, i.e., a value for which the results will tend to. Intuitively, it is the theoretical mean value of a random variable over a large number of experiments, and it is commonly used to summarize all the information about a random variable in a single numerical value.

In Mobile Crowdsourcing scenarios, signal-strength samples are taken by real end-user devices with custom measurement apps. This leads to important sources of error to take into account when aggregating values inside an area:

1.　Measurements are not uniformly distributed in the area, as they are defined by human mobility patterns [19].
2.　The number of measurements in small areas (e.g., coverage area of a single cell) could not be high enough to be considered representative enough [1].
3.　The measurements present accuracy errors in both signal-strength values and geographic coordinates [1].

Hence, some commonly used methods to characterize signal strength could not necessarily return a good estimation of expected value of signal strength, since they do not take into consideration the sources of error aforementioned, which are present in most Mobile Crowdsourcing signal-strength data.

Our paper presents two clear contributions. First, we present a formal analysis about how signal-strength values must be handled to avoid some common pitfalls in using log-scaled signal-strength. Second, we present a novel aggregation method based on interpolation of signal strength (ABOI method). Our proposed method obtained consistently lower RMSE values than other commonly used methods at estimating the expected value of signal strength over an area, in both simulated and real scenarios. Consequently, the ABOI method is demonstrated to be more robust against the existing difficulties of real-world measurements.

The rest of this paper is structured as follows. In Section 2, we discuss the literature on aggregating signal strength for a variety of different purposes. Section 3 presents the physical and mathematical formalities regarding how signal-strength values must be handled when applying mathematical operations. In Section 4, we discuss the most used methods to aggregate signal-strength measurements in Mobile Crowdsourcing scenarios, and their unsuitability when dealing with low number of measurements and nonuniformity in spatial distribution. We also present a novel aggregation method based on interpolation of signal strength, to face these real-world difficulties. Section 5 provides the mathematical foundation for using our proposed method to estimate the expected value of signal strength inside an area of interest. Section 6 indicates that our proposed method obtains better results than other commonly used methods for aggregating signal strength, in both simulated and real scenarios. We conclude in Section 7 that for most Mobile Crowdsourcing scenarios, our proposed model based on interpolation should be preferred over the other methodologies, since it has a better performance even when the number of measurements is low and the spatial distribution of the samples is nonuniform, which is a typical case for real Mobile Crowdsourcing data.

2. Related Work

Due to its large impact on the overall QoS of wireless networks, many research works have focused on characterizing the received signal strength for an area of particular interest. These analyses frequently used all the individual signal-strength samples taken by each mobile device sensing the network.

In this way, some works aggregated several signal-strength samples inside the same area into a unique representative signal-strength value to predict user availability [9], measure the effect of weather conditions in the received signal strength [10,11], analyze network performance [12], measure the impact on signal strength of indoor-outdoor context [13] and find correlations between signal strength and other QoS indicators for mobile networks as network congestion [15], throughput [8] and TCP goodput and latency inside the same geographic area [1]. In this paper, we propose an aggregation method that better estimates

the expected value of signal strength than the methods used in the papers mentioned above, especially when using measurements taken in Mobile Crowdsourcing contexts. Thus, the results and conclusions of these works can be refined by using our proposed model.

It is important to notice that there are other research works that also aggregated several signal-strength samples, although not to find a representative signal-strength value. Some of these works used signal aggregation to perform base transceiver station (BTS) localization [20] or to estimate user location [14,21], mostly based on the RADAR system [22]. These studies are not that related to the problem we are referring to in this paper, which is to aggregate signal strength into a representative value to estimate the mathematical expectation of the signal strength.

There are some works that developed high-resolution coverage maps from Mobile Crowdsourcing signal-strength measurements. These maps were created by plotting each empirical sample on the map [23,24] or by interpolating the signal strength in several uniformly distributed points inside the area of interest using linear interpolation [25], using variations of Kriging method [26,27] or by using Gaussian processes that consider a prior knowledge about theoretical path loss models [28]. These coverage maps are useful for tasks that require highly detailed maps, but when analyzing signal coverage in greater areas, the effectiveness of their fine-grained visualizations will decrease as the resolution of the maps decreases. Positioning all individual samples will greatly increase the clutter in the visualizations, defeating the purpose of providing useful information for the measured areas. Therefore, for these cases, it is also important to take into account the aggregation of signal-strength samples, to be able to generalize their results to maps with lower resolution, where it will be convenient to represent the signal strength in areas by only one representative value. Consequently, the method for signal-strength aggregation proposed in this paper could also be useful for these works.

Some works that employed Mobile Crowdsourcing data discussed the problem of not having uniformly distributed samples in the measured area [26,28] and how the spatial distribution of the samples matches population patterns [29]. This is important, since some researchers that used simulated Mobile Crowdsourcing data to evaluate their methods, implicitly assumed uniform spatial distribution of the samples (as shown in Section 4.1). Uniform distribution is not a realistic measurement scenario, especially when samples are taken by real end-user mobile devices. For a better reliability of this paper's results, we consider in our experiments both uniform distribution of signal-strength samples and distribution based on social network theory [19], which is closer to the spatial distribution present in real Mobile Crowdsourcing measurements.

3. Common Pitfalls in Using Log-Scaled Signal Strength

The use of log-scaled signal-strength values is a widespread methodology for analyzing radio frequency measurements. Signal amplitude could vary very widely, and therefore, it could be difficult to analyze and understand the relationships among different values in the linear watt scale. Hence, using log-scale enhances signal-strength visualizations by improving the display range of axes. The use of log-scaled signal-strength values is also attractive since it can lead to a compression of data, requiring fewer bits of information [30].

Log-scaled signal-strength values could be used in dBm units (decibels with reference to one milliwatt) or in Arbitrary Strength Units (ASU), since ASU values are linearly proportional to the received signal strength in dBm, and consequently, they are also logarithmic values.

Before using and manipulating dBm values, it is important to analyze the origin of dBm from a physical point of view, regarding *dimensional analysis*. Power is a derived quantity that can be expressed in terms of fundamental units (time, length, and mass). In fact, power values must have dimension ML^2T^{-3}. The International System of Units (SI) describes the watt (symbol: W) as a unit of power, defined as a derived unit in terms of base units, where $1\,\text{W} = 1\,\text{kg} \cdot \text{m}^2 \cdot \text{s}^{-3}$. In addition, the prefix *Milli-* (symbol m) has been part of the SI since 1960, and it only denotes a factor of 10^{-3}. This prefix never changes the

units, and therefore, values expressed in milliwatt (mW) are actually being expressed in watt (W) units.

If we consider a power value P_{mW} expressed in milliwatts, its corresponding value in dBm is formulated as

$$P_{dBm} = 10 \log_{10}\left(\frac{P_{mW}}{1 \text{ mW}} \right) \tag{1}$$

It is clear to see that the right side in Equation (1) is dimensionless, since mW dimension is canceled in $\frac{P_{mW}}{1\,\text{mW}}$. Then, the parameter inside the logarithm function in Equation (1) is a dimensionless number, and therefore, P_{dBm} is also dimensionless. This fact is essential because, even when P_{mW} has dimensions, there is no physical sense for P_{dBm} to have it. Thus, it is a mistake to consider dBm as a power unit, since it does not meet the dimension of power quantities (ML^2T^{-3}).

The above is documented by Sonin [31]: "*Products, ratios, powers, and exponential and other functions such as trigonometric functions and **logarithms** are defined for numbers, but **have no physical correspondence in operations involving actual physical quantities**"*.

Furthermore, the mere fact that P_{dBm} is defined as a logarithmic function implies that dBm is a dimensionless quantity. In fact, we consider the formal definition of $10 \log_{10}(x)$:

$$10 \log_{10}(x) = \frac{10}{\ln(10)} \int_{t=1}^{t=x} \frac{dt}{t}$$

The integral $\int_{t=1}^{t=x} \frac{dt}{t}$ corresponds to the sum of an infinite number of terms $\frac{dt}{t}$. All these terms are dimensionless, and therefore, the whole expression $10 \log_{10}(x)$ will always be dimensionless. Then, further interpretations of the dimensionality of dBm are not accepted: dBm values are intrinsically dimensionless.

This non-coherency in dimensionality between dBm values and power quantities (ML^2T^{-3}) can be demonstrated with the knowledge of the use of dimensional formulas in changing units [32], where there is no possible transformation to consistently change from W to dBm. The consequence of the aforementioned is that dBm values do not meet *Bridgman's principle of absolute significance of relative magnitude* (Lemma 1), which is essential to all the systems of measurement in scientific use [33].

Lemma 1. *dBm values **do not** meet Bridgman's principle of absolute significance of relative magnitude.*

Proof. Let SQ be a secondary quantity described by

$$SQ = f(\alpha, \beta, \gamma, ...),$$

where α, β, γ, ... are primary quantities and f is the function that combines them.

SQ satisfies Bridgman's principle of absolute significance of relative magnitude if

$$\frac{f(\alpha_1, \beta_1, \gamma_1, ...)}{f(\alpha_2, \beta_2, \gamma_2, ...)} = \frac{f(x\alpha_1, y\beta_1, z\gamma_1, ...)}{f(x\alpha_2, y\beta_2, z\gamma_2, ...)} \tag{2}$$

holds for all values of α_1, β_1, γ_1, ..., α_2, β_2, γ_2, ... and for all coefficients x, y, z, ... [33].

As stated in Equation (1), dBm can be described as a function of primary quantities:

$$f_{dBm}(\alpha, \beta, \gamma) = 10 \log_{10}\left(\frac{\alpha[kg] \cdot \beta[m^2] \cdot \gamma[s^{-3}]}{10^{-3} \cdot 1[kg] \cdot 1[m^2] \cdot 1[s^{-3}]} \right) \tag{3}$$

Proceeding by contradiction, assume that dBm values do meet Bridgman's principle of absolute significance of relative magnitude. Then, Equation (2) should hold for f_{dBm} in Equation (3) and for all values of $(\alpha_1, \beta_1, \gamma_1)$, $(\alpha_2, \beta_2, \gamma_2)$ and (x, y, z). In particular, it should hold for the following values:

$$
\begin{array}{lll}
\alpha_1 = 10^{-9} & \beta_1 = 1 & \gamma_1 = 1 \\
\alpha_2 = 10^{-8} & \beta_2 = 1 & \gamma_2 = 1 \\
x = 10 & y = 1 & z = 1
\end{array}
$$

By replacing these values in the left side of Equation (2):

$$
\frac{f_{dBm}(\alpha_1, \beta_1, \gamma_1)}{f_{dBm}(\alpha_2, \beta_2, \gamma_2)} = \frac{-60}{-50} = 1.2
$$

And by replacing these values in the right side of Equation (2):

$$
\frac{f_{dBm}(x\alpha_1, y\beta_1, z\gamma_1)}{f_{dBm}(x\alpha_2, y\beta_2, z\gamma_2)} = \frac{-50}{-40} = 1.25
$$

Both sides of the equation are not equal, which is a contradiction. Then, since the relationship is not fulfilled for all values, we conclude that dBm values (represented as f_{dBm}) do not satisfy Bridgman's principle of absolute significance of relative magnitude. □

As a direct consequence of Lemma 1, equations involving dBm units are considered to be not physically relevant [31].

Meeting Bridgman's principle is, according to Percy Bridgman [33], essential to all the systems of measurement in scientific use. This principle is fundamental to guarantee that the selection of a different unit of measurement will not affect the outcomes of any experiment. Therefore, as dBm values do not meet Bridgman's principle, some numerical relationships among power values in W do not remain true when using dBm, i.e., the outcomes of scientific experiments can be affected if using dBm values instead of watt values. This should not be allowed in scientific research, as nature is indifferent to the arbitrary choices we make when we pick base units. Indeed, as Sonin [31] precisely stated: "*Nature is indifferent to the arbitrary choices we make when we pick base units. We are interested, therefore, only in numerical relationships that remain true independent of base unit size.* ". However, dBm values do not respect this, as demonstrated in the following straightforward example:

$$
1 \text{ mW} + 1 \text{ mW} = 2 \text{ mW}
$$

If we transform all values from mW to dBm using Equation (1), we have

$$
0 \text{ dBm} + 0 \text{ dBm} = 3.0102999566 \text{ dBm}
$$

This is, of course, wrong and contradictory, and it exemplifies that if we wrongly attempt to perform addition of dBm values, we will reach erroneous conclusions such as 1 mW = 2 mW. Accordingly, to be coherent with the dimensional analysis and with the mathematical basis, all mathematical operations involving signal strength must use linear watt values.

Some research from before the 1980s purposely included these wrong methodologies in their analyses. However, when applying mathematical operations to log-scaled signal values, they had a clear understanding of the definition and implications of using logarithmic power values. As they stated, they performed these methodologies to compare how different their results would be if using log-scaled signal values [34,35], or to explore the *"attractiveness of the logarithm of power"* such as its contribution to a compression of data requiring fewer bits of information [30] (what may have been a real concern at that time). Nevertheless, we did not find any discussion or argument on why to use log-scaled signal values in more recent papers. Indeed, many of these papers manipulated dBm values without mentioning the correspondence between dBm and watt values [13,15,20,21,27,36–41], and moreover, some of them manipulated signal-strength values without reporting the unit of measurement employed [12,14,29,42,43]. Many of the papers that followed these wrong methodologies got log-scaled signal-strength measurements directly from mobile operating systems (Android or iOS) [1,8,11–13,15,20,23,26,27,36–38,40]. Therefore, it is

plausible that they just used and manipulated the data returned by the systems without a thorough analysis about the unit of the collected signal-strength values.

It is important to understand that applying mathematical operations with log-scaled signal-strength values involves wrong models and interpretations of reality, and therefore, leads to wrong conclusions. Nevertheless, related research works have frequently made these mistakes. Many of these papers have been published during the last few years, demonstrating that the misuse of log-scaled signal-strength values is a real problem within the mobile computing community presently. These methodologies must be avoided, even when they have been constantly used in the past, since their habitual use is not a valid argument against their contradiction with some basic principles of scientific analysis.

The following subsections describe some of the common, but misinterpreted practices.

3.1. Averaging Signal Strength

The average of signal-strength measurements taken in similar temporal space conditions has been widely employed. For instance, the arithmetic mean of measurements taken in a single point can be used to reduce measurement variance, since every signal-strength sample is assumed to be contaminated with unrelated additive noise. Moreover, the arithmetic mean of measurements inside the same geographic area can be used to obtain a representative value of signal strength, getting an estimation of the mathematical expectation of signal strength in the area (As shown in Section 4.1).

The arithmetic mean involves taking the sum of samples; however, we already stated the lack of physical sense and relevance of the addition (and any other equation) involving log-scaled signal values. Consequently, the arithmetic mean of log-scaled signal values cannot either be considered to be physically relevant. This bad practice implies in most cases a distortion of real signal-strength behavior [30,35], as shown in the following simple but explanatory example. Let a be a vector of signal-strength values in dBm units:

$$a = \begin{bmatrix} -45 \text{ dBm} & -55 \text{ dBm} \end{bmatrix} \tag{4}$$

The arithmetic mean of samples in a (in linear scale) is 1.74e−5 mW, which is equal to −47.6 dBm. Instead, the arithmetic mean of log-scaled samples in a is −50 dBm, with an error of 2.4 dB from the real value introduced by this misleading methodology. Although these errors may seem small in some cases, they should not be underestimated due to the impact of signal-strength fluctuations on other important network performance metrics [1–4]. Differences around 5 dB in signal strength could imply in some cases an increase of 100% in packet loss rate and round-trip time of a connection over the mobile network [3].

Despite the aforementioned, there are works in which several signal-strength samples were aggregated by performing a log averaging process, taking the arithmetic mean of dBm or ASU measurements, misunderstanding signal-strength real behavior. Some works that used this incorrect methodology are listed below:

- (2017) Sabu et al. [11] conducted a correlation study between signal strength and rainfall intensity in an area of interest, where logarithmic ASU values were aggregated by taking the arithmetic mean. As result, the authors concluded that the drop of signal strength during rainfall was not as significant as expected by the theoretical hypothesis.
- (2018) In the data exploration section provided by Sung et al. [8], an area of interest was divided in smaller square areas. For each square area, signal strength was reported by taking the mean of several logarithmic ASU measurements. As result, a weak geographical correlation between signal strength and throughput was found.
- (2015) In the research work of Marina et al. [13], signal-strength samples in dBm units were divided according to their context (indoor or outdoor), and then aggregated by taking the arithmetic mean. Then, the authors analyzed the great impact of user context (indoor or outdoor) on the received signal strength.

- (2013) Sonntag et al. [1] created signal-strength coverage maps by taking the arithmetic mean of values represented as percentages. The percentage values are calculated linearly from logarithmic signal strength, so they are also logarithmic values. The authors concluded that coverage maps created from crowdsourced signal strength were not very good at presenting the actual transport quality.

In addition to the above, there are other research papers that also followed the incorrect methodology of calculating the arithmetic mean of logarithmic signal-strength values [12,17,20,21,36–38,40].

The only recorded case in radio frequency analysis where directly average log-scaled measurements using the arithmetic mean is admitted, is when dealing with samples of a repetitive or continuous wave signal. For this case, averaging log-scaled values is equal to log-scaling the average of linear values [44]. However, this is not the case of the signals we are referring to in this paper. Therefore, the introduced error by averaging log-scaled values depends on the statistics of the power estimates being averaged [30,34]. Consequently, considering that log-scaling the linear average will be equal to the average of the log-scaled values, and that the introduced error can be ignored, are incorrect assumptions.

3.2. Comparing Signal Strength

When comparing two signal-strength values, e.g., to calculate a prediction error, it is essential to properly measure the difference between the values. For example, if the signal strengths we want to compare are -50 dBm and -45 dBm, then it is correct to say that the values differ in 5 dB, which represents the relation between both signal strengths. It is also correct to say that the difference between these values is 2.162e-5 mW (the difference in linear scale) or, equivalently, -46.65 dBm, although this latter form in dBm units can be quite confusing. However, it is a big mistake to say that the difference between -50 dBm and -45 dBm is 5 dBm, since this is equal to 3.162 mW, which is several orders of magnitude greater than the real difference shown before. Despite the above, some works performed signal-strength comparisons using this last incorrect methodology [13,36,40]. This misunderstanding is demonstrated in sentences as *"median error is 6 dBm"* [40], *"differing by more than 15 dBm"* [13], *"the real signal strength is 2.5 dBm stronger"* [38] or *"the errors are 10 dBm, 7 dBm, and 6 dBm, respectively"* [36]. These error levels are not coherent with the data used, where there are practically no values greater than -50 dBm.

In addition, when comparing different signal-strength prediction errors to decide which error is lower, it is very important to compare them using absolute errors (in watt or dBm scale) rather than relative errors (in dB). The comparison of prediction errors using relative values, could lead to misunderstandings illustrated in the following simple but explanatory example. An error of $+4$ dB at predicting -70 dBm could be considered lower than an error of $+5$ dB at predicting -110 dBm, holding that 4 dB $<$ 5 dB. However, if we analyze these errors as linear absolute errors, we find that the first error is actually more than 10,000 times greater than the second one, giving an absolutely opposite view than the obtained by analyzing relative dB values. An example of the use of this incorrect methodology is described in the following:

- (2019) Recently, Alimpertis et al. [45] proposed a new method based on machine learning to perform signal-strength prediction, i.e., given a set of signal-strength measurements in an area, estimate signal-strength values in other singular points. They claimed that their method consistently obtains lower prediction errors than related state-of-art algorithms. Nevertheless, it can be shown that the comparison of errors used by them leads to inconclusive results. Using the values shown in Table 4 of [45], for cell ID *x*204, their method obtains an average error of 2.3 dB, outperforming Ordinary Kriging (OK) and Ordinary Kriging Detrending method (OKD) which obtains average errors of 3.85 dB and 2.99 dB respectively. However, if we consider the case in which their method's error is $+2.3$ dB, OK's error is -3.85 dB, and OKD's error is -2.99 dB, all in relation to the expectation of the signal strength

of cellID $x204$ (-96 dBm), we have that OK method has an error 26% lower than their method's error, and OKD method has an error 30% lower than their method's error (using linear watt scale). In that case (a possible case given the prediction errors stated in the paper), their method actually gets worse results than related state-of-art algorithms.

Another common task is to summarize several prediction errors for $(real\,forecast)$ pairs of signal-strength values. This is performed by calculating different measures of prediction accuracy as MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), or MASE (Mean Absolute Scaled Error). Nevertheless, many of these works failed at estimating their prediction accuracy by using a logarithmic scale for prediction errors, and summarizing them by applying mean-based aggregation as MAPE [26,27], MSE (Mean Squared Error) [28], or RMSE (Root Mean Squared Error) [45]. Then, they added a source of error at applying mean functions to log-scaled values, as mentioned in Section 3.1. In fact, since it is well known that the average of the logs will always be less than or equal to the log of the average [35], applying mean-based error measures to log-scaled errors, will imply an underestimation of real errors.

Signal-strength samples in linear scale should be preferred for estimating the error between two signal-strength values and for summarize several errors in an accuracy measure. However, this does not prevent these results from being used latter in log-scale if desired (for example, for visualization).

4. Signal-Strength Aggregation

As mentioned in Section 1, we consider each aggregated value as an estimation of the mathematical expectation of signal strength:

For an area A, we consider the function $P(\vec{p})$, which represents the signal strength in function of the position $\vec{p}$. Thus, a representative signal-strength value for A, obtained from the aggregation of individual measurements, will try to be as close as possible to the mathematical expectation $\mathbb{E}(P(X))$, where X is a uniformly distributed random variable of position in A.

We define $\overline{P}_A$ as the division of the integral of the function P in A, and the total area A, which is equal to $\mathbb{E}(P(X))$ as shown below:

$$\mathbb{E}(P(X)) = \int_{\Omega} P(\omega) f_X(\omega) d\omega = \frac{\iint_A P(\vec{p}) dA}{\iint_A dA} =: \overline{P}_A \tag{5}$$

where Ω denotes the set of all positions ω in A, f_X is the probability density function of X, which is a constant equal to $\frac{1}{\iint_A dA}$, and $\vec{p}$ denotes the position in A.

Considering a discretization of the space, the mathematical expectation in Equation (5) can be approximated by Riemann sums:

$$\overline{P}_A \approx \frac{\sum_{i=1}^{m} P(x_i) \Delta A_i}{\sum_{i=1}^{m} \Delta A_i} \tag{6}$$

and, when the discretization is such that all the points are equispaced, it follows that $\Delta A_i = \Delta A$ is constant for all $i = 1, ..., m$. The approximation becomes better as ΔA gets smaller, and consequently, the number of points (denoted by $m_{(\Delta A)}$) gets larger. Accordingly, the approximation by Riemann sums corresponds to the arithmetic mean and fulfills that:

$$\lim_{\Delta A \to 0} \frac{1}{m_{(\Delta A)}} \sum_{i=1}^{m_{(\Delta A)}} P(x_i) = \overline{P}_A \tag{7}$$

Thus, as the equispaced discretization becomes finer, the better the approximation of $\overline{P}_A$. The main drawback for this approximation method is the need to know the value of P in several equispaced positions over A to obtain an accurate estimation.

Another strategy to approximate the expected value relies in considering that the positions of the measurements x_i, $i = 1, ..., m$, are given by independent uniform random variables over the area A. Then, we can use the Monte Carlo method to approximate $\overline{P}_A$ and, by the law of large numbers, we have that

$$\lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} P(x_i) = \overline{P}_A \tag{8}$$

The expression in Equation (8) is equivalent to Equation (7), and corresponds to the arithmetic mean. This is the value that past works referred to as the "local mean signal strength", used to summarize signal strength in areas of a few meters (up to 40 wavelengths) [16–18]. In fact, local mean signal strength *"is obtained by averaging a large number of individual RF measurements taken in a local neighborhood"* [16]. Thus, related studies that estimate local mean signal strength are actually estimating the mathematical expectation of signal strength.

In the following, we discuss algorithms for estimating $\overline{P}_A$, using data from Mobile Crowdsourcing apps. In the first place, we consider as aggregation methods two commonly used performance metrics: arithmetic mean and median value. In addition, we propose a novel method based on the interpolation of signal-strength values.

4.1. Arithmetic Mean

A simple method to summarize signal-strength measurements is to take the arithmetic mean $\overline{x}_A$ of all the samples in area A, as commonly used in Mobile Crowdsourcing contexts [1,8,11,13,16,29]:

$$\overline{x}_A = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{9}$$

This is a good first approach to estimate $\overline{P}_A$, based on the fact that if the samples are independent uniformly distributed in area A, then $\overline{x}_A$ is an example of Monte Carlo method, which assures that $\overline{x}_A$ converges to $\overline{P}_A$ when $n \to \infty$, as shown in Equation (8). Indeed, the law of large numbers and the Monte Carlo method could apparently justify the use of the arithmetic mean as an estimator of the mathematical expectation of signal strength in area A. Nevertheless, research studies that used the arithmetic mean over signal-strength samples did not look over the fulfillment of the hypothesis required by the Monte Carlo method. First, in a realistic Mobile Crowdsourcing scenario, the number of signal-strength samples could be low for small areas. Second, crowdsourced signal-strength measurements would not be sampled uniformly on area A, as their positions are determined by human mobility patterns. Accordingly, there is no real mathematical foundation for using the arithmetic mean as estimator of the expected value of signal strength in these measurement contexts.

Estimating $\overline{P}_A$ by taking the arithmetic mean of Mobile Crowdsourcing data, is based on a *convenience sampling* process that only considers measurements from locations that are readily available or easy to reach. Readily available locations are directly defined by the mobility of test users. The use of this sampling method is well known to be likely to have biased results, because selecting cases based on their availability does not allow a generalization to the total population [46]. In our case, this means that the estimation of $\overline{P}_A$ will be biased by the locations in area A where test users took measurements.

As mentioned in Section 3.1, the arithmetic mean of signal samples should be taken over linear values to avoid induced bias due to incorrect methodologies. In addition to these physical and mathematical formalities, the importance of using linear values to better estimate the expected value of signal strength has also been stated in the past: *"In terms of*

*accuracy, the preferred method for estimating the local mean signal strength at a specific point is to average (in **watts**) a **large number** of individual RF measurements"* [16]. Nevertheless, some works use the wrong methodology by explicitly applying the arithmetic mean over signal strength in logarithmic scale [1,8,11,13] as stated in Section 3.1.

Due to the above, in this paper we consider only the correct mean of samples in linear scale.

4.2. Median Value

Another method to summarize several signal-strength samples is by choosing the median value that separates the higher half from the lower half of all the measurements [10,47–49]. The idea behind using median value to aggregate signal-strength samples is that it is not skewed so much by a small proportion of extremely large or small values, which is a common situation in this case study, because, as shown in Section 3, signal amplitude could vary very widely among measurements.

Furthermore, since the logarithm is a strictly increasing function, the median value has the advantage that it will be the same value if selected from signal-strength values in linear (watt) or logarithmic (dBm) scale.

In cases as the Gaussian distribution, the median value is a good estimator for the mathematical expectation, since the latter naturally separates the higher from the lower half of possible values. Nevertheless, this assumption is not very likely to be true for Mobile Crowdsourcing contexts, where the signal-strength distribution depends on the positions of base transceiver stations with respect to the area of interest.

In addition to the above, the median value method also induces a bias due to the *convenience sampling* of measurements. Therefore, the median value is not expected to perform well at estimating the mathematical expectation of signal strength in Mobile Crowdsourcing scenarios, as there is no mathematical foundation for its use. However, due to its wide use in the literature, it is important to consider the median value as a baseline of our study to quantify the error it can reach at estimating $\overline{P}_A$.

4.3. Our Proposal: Average Based on Interpolation (ABOI Method)

As shown in Section 3.1, most Mobile Crowdsourcing scenarios do not fulfill the required hypotheses to employ the arithmetic mean as an estimator of the mathematical expectation of signal strength (hypotheses for Monte Carlo integration). Therefore, we wanted to design a more robust method to estimate the expected value from signal-strength measurements, without requiring the samples to be independent and uniformly distributed.

For our proposed method, we return to the idea of estimating the mathematical expectation of signal strength using Riemann sums, according to Equation (6). As mentioned in Section 1, to obtain better approximations to the real $\overline{P}_A$ by using Riemann sums, we need an equispaced grid of signal-strength values as fine-grained as possible. However, as stated in Section 2, it is not possible to ensure high number of measurements and uniform spatial distribution in most Mobile Crowdsourcing scenarios. To solve these problems, we use the available measurements to interpolate the signal strength in a fine-grained grid, obtaining equispaced data and increasing the number of available samples. Thus, to estimate $\overline{P}_A$ we take the arithmetic mean of all values in the fine-grained grid G (in watt) as shown in Equation (7), avoiding the difficulties of nonuniform spatial distribution and low number of measurements. Consequently, to obtain a good estimation of the mathematical expectation, we need to establish the conditions on the signal-strength measurements that guarantee a proper interpolation. As our proposed method is an average based on interpolation, we will refer to it as ABOI.

Although there are many interpolation methods, it is out of the scope of this paper discussing the advantages and disadvantages of each one. For the interpolation step in the ABOI method, we use one of the simplest and commonly employed interpolations methods in signal-strength analysis: the Ordinary Kriging (OK) algorithm. Nevertheless, the ABOI method could be improved by using a more complex and accurate interpolation algorithm.

To estimate the value of signal strength at a position x_0 on the grid, the OK algorithm takes a linear combination of its neighbors:

$$P^*(x_0) = \sum_{i=1}^{n} \omega_i P(x_i) \tag{10}$$

where x_i represents all the neighbors of x_0, and ω_i is the corresponding weight of each neighbor. In general, ω_i is proportional to the distance between x_0 and x_i.

It is important to notice that many authors wrongly used this method with dBm values or simply do not make explicit the scale used [36,41,43,45]. As mentioned in Section 3, we emphasize that this algorithm should be used on the linear power scale, since it involves algebraic operations, such as addition and weighting.

5. Mathematical Foundation for the Use of the ABOI Method

In this Section, we present the mathematical foundation for using the ABOI method to estimate the expected value $\overline{P}_A$ of signal-strength measurements inside an area A of interest. First, we announce Theorem 2, establishing the conditions under which the error of the estimation of $\overline{P}_A$ provided by the ABOI method can be smaller than ε. The hypotheses required for this result are shown to be consistent with realistic Mobile Crowdsourcing scenarios, contrarily to the case of arithmetic mean, as stated in Section 4.1. Lastly, we demonstrate that the ABOI method is an improvement on arithmetic mean at estimating $\overline{P}_A$, i.e., signal-strength measurements that are favorable for the arithmetic mean (that do fulfill Monte Carlo integration hypotheses) are still favorable for the ABOI method. However, favorable cases for the ABOI method can be very disadvantageous for the arithmetic mean.

5.1. ABOI Theorem

ABOI Theorem (Theorem 2) specifies the conditions under which the error of ABOI's estimation can be smaller than ε, providing a proper approximation of the expected value of signal strength. For that purpose, some important definitions need to be stated first.

Let $N = \{x_1, x_2, ..., x_n\}$ be the set of positions of the initial n signal-strength measurements taken inside a rectangle area $A = [a_1, b_1] \times [a_2, b_2]$. Analogously, let M be the set of positions of the m points equispaced over A on which the ABOI method interpolates signal strength. Sets N and M are exemplified in Figure 1.

Let us consider the following definition of the fill-distance:

$$h_N := \sup_{x \in A} \min_{x_i \in N} ||x - x_i|| \tag{11}$$

The value h_N indicates the largest distance between each position in A and its nearest neighbor in N (original measurements).

Let us call ABOI(N,M) the return value of the ABOI method after using the n original measurements to interpolate signal strength (using OK) on grid M, and computing the arithmetic mean of the m interpolated watt values.
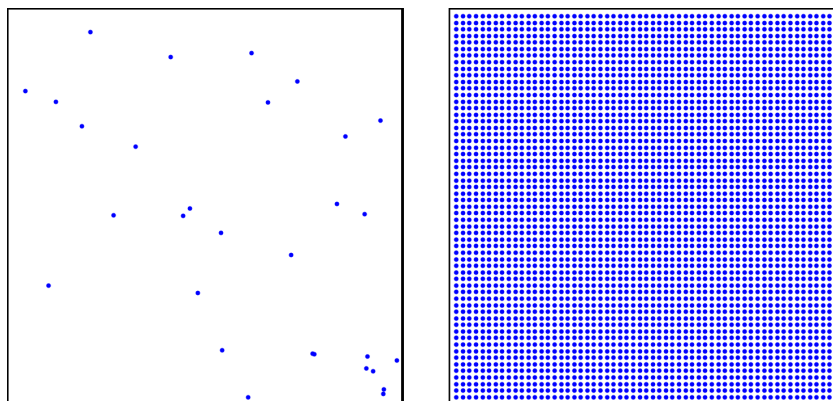
Figure 1. Example of set N with $n = 30$ positions of initial measurements (**left**), and set M with $m = 3481$ equispaced positions over A where to interpolate signal strength (**right**).

Theorem 2. *Given that power measurements $P(\cdot)$ can be modeled by Gaussian Processes ($\star$). Let $\varepsilon > 0$ be the desired error level for the estimation of the expected value provided by ABOI(N)(M). Let A be a rectangle area where to estimate the mathematical expectation of signal strength. If the n initial measurements are such that h_N is small enough ($\star\star$), and selecting M as a fine-grained enough grid over A ($\star\star\star$), then*

$$\mathbb{E}\left[\left|\mathbb{E}_A(P(X)) - \text{ABOI(N,M)}\right|\right] \leq \varepsilon$$

that is, the expected value of the error between the mathematical expectation of signal strength over A and the estimation provided by the ABOI method is smaller than the given ε.

Proof. Let P_M be the arithmetic mean of real signal-strength values on each position in M. These m values are not known when applying the ABOI method. Nonetheless, P_M will be helpful to bound the expected value of the estimation error. Indeed, we can bound the estimation error of the ABOI method as follows:

$$\left|\mathbb{E}_A(P(X)) - \text{ABOI(N,M)}\right| = \left|\mathbb{E}_A(P(X)) - P_M + P_M - \text{ABOI(N,M)}\right|$$
$$\leq \underbrace{\left|\mathbb{E}_A(P(X)) - P_M\right|}_{(I)} + \underbrace{\left|P_M - \text{ABOI(N,M)}\right|}_{(II)} \tag{12}$$

We will bound (I) and (II) separately.

$(I) : \left|\mathbb{E}_A(P(X)) - P_M\right|$

As stated in Equation (7), we can define $\mathbb{E}_A(P(X))$ as follows:

$$\mathbb{E}_A(P(X)) = \lim_{\Delta A \to 0} \frac{1}{m_{(\Delta A)}} \sum_{i=1}^{m_{(\Delta A)}} P(x_i) \tag{13}$$

or equivalently, for all $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\Delta A \leq \delta \implies \left|\mathbb{E}_A(P(X)) - \frac{1}{m_{(\Delta A)}} \sum_{i=1}^{m_{(\Delta A)}} P(x_i)\right| < \varepsilon/2 \tag{14}$$

where $\dfrac{1}{m_{(\Delta A)}} \displaystyle\sum_{i=1}^{m_{(\Delta A)}} P(x_i)$ is analogous to what we previously defined as P_M.

Therefore, hypothesis ($\star\star\star$) allows us to select a fine-grained enough grid M that gives us the desired error bound $\varepsilon/2$.

Selecting M as aforementioned, we have that

$$\left| \mathbb{E}_A(P(X)) - P_M \right| \leq \varepsilon/2 \tag{15}$$

$(II):\left| P_M - \text{ABOI(N,M)} \right|$

As previously mentioned, P_M is the arithmetic mean of real signal strength over M (m unknown values), whereas ABOI(N,M) is the arithmetic mean of the m interpolated values on grid M obtained by OK interpolation of the original n measurements in N. Therefore, P_M and ABOI(N,M) are defined as follows:

$$P_M = \left(\sum_{i=1}^{m} P(x_i) \right) / m$$

$$\text{ABOI(N,M)} = \left(\sum_{i=1}^{m} I_N(x_i) \right) / m$$

where $P(\cdot)$ corresponds to the real signal strength, and $I_N(\cdot)$ corresponds to the OK interpolation of the original n signal-strength measurements in N. Accordingly,

$$
\begin{aligned}
\left| P_M - \text{ABOI(N,M)} \right| &= \left| \left(\sum_{i=1}^{m} P(x_i) \right) / m - \left(\sum_{i=1}^{m} I_N(x_i) \right) / m \right| \\
&= \left| \left(\sum_{i=1}^{m} P(x_i) - I_N(x_i) \right) / m \right| \\
&\leq \left(\sum_{i=1}^{m} \left| P(x_i) - I_N(x_i) \right| \right) / m \\
&\leq \max_{i \in [1:m]} \left| P(x_i) - I_N(x_i) \right|
\end{aligned}
$$

Thus, the difference between P_M and ABOI(N,M) is bounded by the maximum interpolation error among all the m positions of grid M. Wang et al. [50] provided an exhaustive analysis regarding this maximum interpolation error of OK. Indeed, hypothesis ($\star$) allows the use of Corollary 1 of Wang et al. [50] along with Theorem 11.22 of Wendland [51] to obtain the following result (a detailed description of this outcome is provided in Appendix A):

$$\lim_{h_N \to 0} \mathbb{E}\left[\max_{i \in [1:m]} \left| P(x_i) - I_N(x_i) \right| \right] = 0 \tag{16}$$

or equivalently, for all $\varepsilon > 0$, there exists a $\bar{h}$ such that

$$h_N \leq \bar{h} \implies \mathbb{E}\left[\max_{i \in [1:m]} \left| P(x_i) - I_N(x_i) \right| \right] \leq \varepsilon/2$$

Therefore, hypothesis ($\star\star$) gives us the conditions such that h_N is small enough to guarantee the desired error bound $\varepsilon/2$:

$$\mathbb{E}\left[\left| P_M - \text{ABOI(N,M)} \right| \right] \leq \varepsilon/2 \tag{17}$$

Finally, by joining the bounds for (I) and (II), i.e., by plugging (15) and (17) into (12), we obtain the desired inequality

$$\mathbb{E}\left[\left| \mathbb{E}_A(P(X)) - \text{ABOI(N,M)} \right| \right] \leq \varepsilon$$

which completes the proof. □

5.2. Improvement on Arithmetic Mean

In this Section, we show that the ABOI method is an improvement on arithmetic mean at estimating $\overline{P}_A$. Both methods require specific conditions about the number and position of the initial signal-strength measurements. On the one hand, ABOI requires h_N to be small enough (hypothesis $(\star\star)$ of Theorem 2). On the other hand, the arithmetic mean requires the initial measurements to fulfill Monte Carlo integration hypotheses. Only if satisfying these conditions, the methods can be considered to be appropriate for estimating the mathematical expectation of signal strength. In the following, we will show that:

Preposition 2.1. *If the initial measurements allow the arithmetic mean to be considered to be an appropriate estimator of $\overline{P}_A$, this implies that the ABOI method will also be considered to be an appropriate estimator of $\overline{P}_A$.*

Preposition 2.2. *If the initial measurements allow the ABOI method to be considered to be an appropriate estimator of $\overline{P}_A$, this does not imply that the arithmetic mean will be considered to be an appropriate estimator of $\overline{P}_A$.*

Proof of Preposition 2.1. If the initial conditions allow the arithmetic mean to be considered to be an appropriate estimator of $\overline{P}_A$, then the set of n signal-strength measurements fulfill Monte Carlo integration hypotheses (Section 4.1), i.e., the number n of measurements is high enough, and they are independent and uniformly distributed over the area. Theorem 6.6 of Niederreiter [52] suggests a bound for h_N derived from its (extreme) discrepancy $D_n(N)$,

$$h_N \leq \sqrt{2} D_n^{1/2}(N)$$

where N is the set of positions of the n initial measurements. Given that the positions in N are independent random variables uniformly distributed over the area, Pronzato [53] states that

$$D_n(N) = \mathcal{O}[(\log n)^2/n]$$

This result indicates that after a given number of measurements, hypothesis $(\star\star)$ will be satisfied. Therefore, the ABOI method will also be considered to be an appropriate estimator of $\overline{P}_A$. □

The intuition behind Preposition 2.1 is that in case of measurements uniformly distributed over the area, both methods can be considered to be appropriate to estimate $\overline{P}_A$. However, as discussed in Section 1, uniform spatial distribution is an unrealistically optimistic case for crowdsourced measurements.

Proof of Preposition 2.2. If the conditions allow the ABOI method to be considered to be an appropriate estimator of $\overline{P}_A$, then hypothesis $(\star\star)$ is fulfilled. This hypothesis only requires h_N to be small enough and does not require any specific distribution of measurements over the area. In particular, it does not require the measurements to be independent nor uniformly distributed over the area, which are necessary conditions for fulfilling Monte Carlo integration hypotheses. Therefore, the arithmetic mean may not be considered to be an appropriate estimator of $\overline{P}_A$. □

The intuition behind Preposition 2.2 is that the requirements of ABOI are less restrictive and more likely to be true in Mobile Crowdsourcing scenarios. As mentioned before, uniform spatial distribution is not a realistic case for measurements taken by real users, and therefore, there is no mathematical foundation for using the arithmetic mean (Section 4.1). However, real crowdsourced data is still able to fulfill hypothesis $(\star\star)$, so far as the number of measurements allows it. Indeed, it is certainly expected that if the number of measurements is very low, then ABOI's estimation will not be accurate, since h_N will hardly be small enough. Likewise, if the number of signal-strength measurements is high,

then ABOI's estimation will be inclined to be closer to $\overline{P}_A$. This is a recurrent condition when estimating values of random effect models from measurements, and therefore, it cannot be avoided due to the stochastic behavior of the observations.

6. Experimental Results

Given that Section 5 provided the mathematical foundations for using the ABOI method to estimate the expected value of signal strength, we wanted to analyze experimentally its suitableness for this task. Additionally, we wanted to compare ABOI against the other aggregation methods commonly employed to estimate the mathematical expectation of signal strength (Section 4). In particular, we were interested in comparing ABOI with the arithmetic mean, as Section 5.2 gives us the intuition that the estimations provided by the ABOI method should be at least as good as the estimations provided by the arithmetic mean.

To evaluate and compare the aggregation methods described in Section 4, we performed experiments in both simulated and real scenarios. As this paper is the first attempt to challenge existing assumptions about signal-strength aggregation, we performed the following simplifications to the problem of estimating the mathematical expectation of signal strength in an area:

1. We considered areas with signal strength coming from only one base transceiver station (BTS).
2. Even when there may be a time variability of signal strength in the area [10,28], we considered that the mathematical expectation is estimated for a static power configuration of the BTS.

6.1. Simulated Scenario

We considered an area A of 500 m × 500 m where a 30-meters tall BTS is placed at the center. We simulated the real signal strength on a fine-grained grid G over A with 5 m spacing, considering long-term attenuation due to path loss equation and medium-term variation due to shadowing modeled by a full covariance matrix [28,54–57]. Indeed, the real signal strength in G is given by

$$\vec{1}P - 10\alpha \log_{10}(\vec{d}) + \vec{v} \tag{18}$$

where $\vec{1}P = [P, P, ..., P]^T$ is a vector with n repeated values of P, the power transmitted by the simulated BTS; α corresponds to the path loss exponent; and $10\alpha \log_{10}(\vec{d})$ is the path loss attenuation, where $\vec{d} = [d_1, d_2, ..., d_n]^T$ is the vector of distances between the position of each measurement and the position of the BTS. In addition, $\vec{v}$ is an attenuation factor due to shadowing effects, where

$$\vec{v} \sim \mathcal{N}(0, \Sigma_v)$$

and the covariance matrix Σ_v is composed of elements given by $\text{Cov}(x_i, x_j) = \sigma_v^2(-d_{ij}/D_{corr})$, where d_{ij} is the distance between the positions x_i and x_j in G, and D_{corr} is a parameter that models the correlation among the measurements.

Next, we simulated signal-strength measurements as if they were taken by real mobile devices, i.e., measurements included long-term attenuation due to path loss equation and medium-term variation due to shadowing, but they also included accuracy errors in both signal-strength values and geographic coordinates (due to hardware inaccuracy). The simulated measurements are given by

$$\vec{X} = \vec{1}P - 10\alpha \log_{10}(\vec{d}) + \vec{u} + \vec{v} + \vec{w} \tag{19}$$

where $\vec{X} = [x_1, x_2, ..., x_n]^T$ is an $n \times 1$ vector that contains the measurements. As in Equation (18), $\vec{v}$ is the attenuation factor due to shadowing effects. Additionally, as geolocation

sensors are not perfectly accurate, position errors are considered when estimating the position of each measurement. This component is simulated by

$$\vec{u} \sim \mathcal{N}(0, \rho_u^2 D),$$

which corresponds to a Gaussian distribution with a mean vector $\vec{0}$ and covariance matrix $\rho_u^2 D$, where $D = \text{diag}\{1/d_1, 1/d_2, ..., 1/d_n\}$ [28]. Finally, $\vec{w}$ in Equation (19) is some unrelated additive noise, where

$$\vec{w} \sim \mathcal{N}(0, \sigma_w^2 I_n).$$

For this simulation, the following values were used: $P = -10$ dBm, $\alpha = 3.5$, $\sigma_w = \sqrt{7}$ dB, $\sigma_v = \sqrt{10}$ dB, $\rho_u = 0.2$ dB and $D_{corr} = 50$ m. This setting is the same used by Santos et al. [28].

Thus, signal-strength values simulated over grid G using Equation (19) generate the spatial field shown in Figure 2.
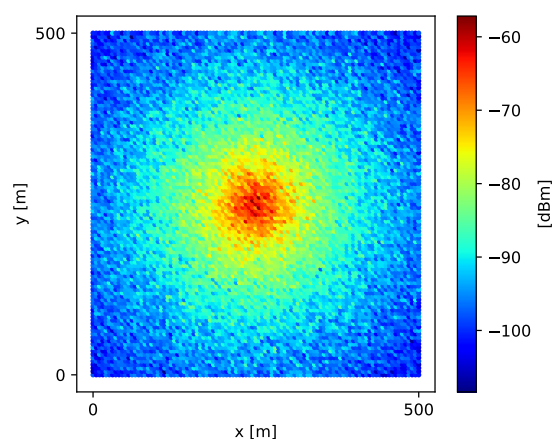
Figure 2. Simulated spatial field of signal strength over a fine-grained grid G.

We calculated the ground truth $\overline{P}_A$ (expected value of signal strength over A) as a Riemann sum considering all values in G.

It is important to clarify that although the simulation model and its parameters were defined using dBm values, we always carefully manipulated signal-strength values using the linear watt scale. Thus, we avoided the mishandling of log-scaled signal-strength values, as discussed in Section 3.

For this experiment, we took different signal-strength measurement sets of sizes 50, 100, 200, 400, 700, and 1000. We distributed the samples on the grid by using two different methods:

1. Completely uniform distribution on the grid, which is commonly used, but not realistic for Mobile Crowdsourcing scenarios, as discussed in Section 1.
2. Considering the mobility model based on social network theory proposed by Musolesi et al. [19]. This model is closer to the spatial distribution of Mobile Crowdsourcing measurements, as they are defined by human mobility.

Figure 3 shows the difference in spatial distribution of 100 samples at using the two methods explained above.
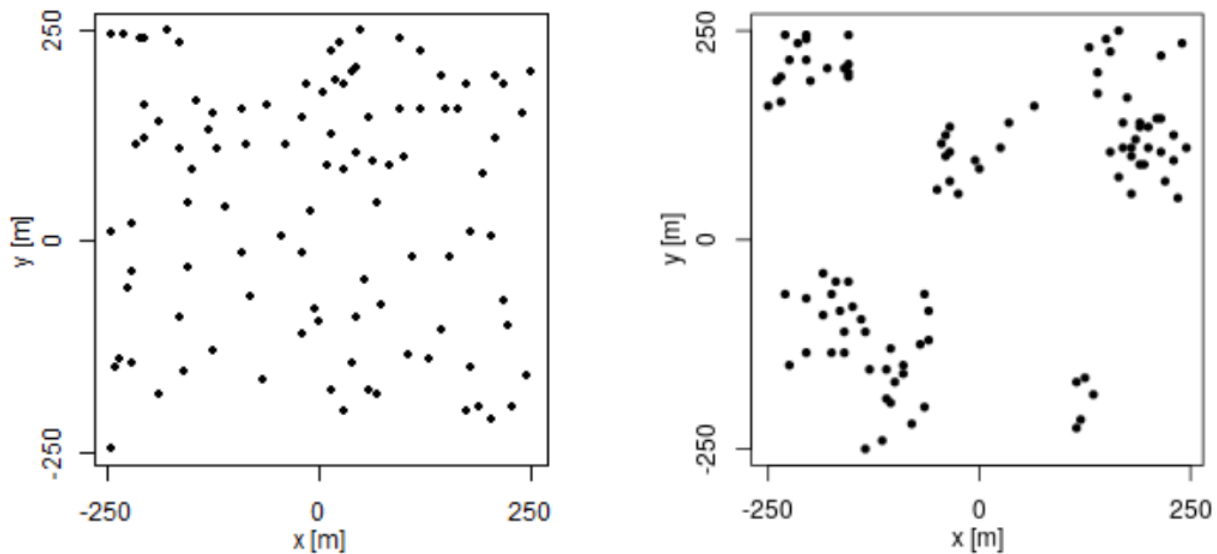
Figure 3. Example of spatial distribution for 100 samples using uniform distribution (**left**) and distribution based on social network theory (**right**).

For every sample size and type of spatial distribution, we estimated $\overline{P}_A$ by applying the three methods presented in Section 4. We repeated each experiment 40 times, i.e., we took 40 different sample sets in every case.

The results for experiments using uniform distribution are shown in Figure 4, where for each aggregation method and sample size we have the boxplot that depict the estimations of $\overline{P}_A$. It is important to clarify that all figures were calculated in linear scale, avoiding the errors mentioned in Section 3. All signal-strength values are shown in pW units, where $1 \text{ pW} = 1 \times 10^{-12}$ W.

As expected, arithmetic mean estimations tended to be close to $\overline{P}_A$, since uniform distribution is its best case, as explained in Section 4.1. Median value performed poorly, predicting nearly constant values far from the real one. Our proposed ABOI method showed satisfactory results and a similar behavior to the arithmetic mean.

In addition, Figure 5 shows the RMSE measure obtained by the aggregation methods, properly calculated using the linear values of the estimations of $\overline{P}_A$, as discussed in Section 3.2. RMSE values for our proposed ABOI rapidly decreased to low values, obtaining very similar results than the arithmetic mean.

These results agree with the intuition of Preposition 2.1, as in the case of uniform spatial distribution, both ABOI and the arithmetic mean performed well at estimating $\overline{P}_A$.

The results for experiments using spatial distribution based on social network theory are shown in Figure 6. The arithmetic mean showed a more erratic performance than before, without a clear convergence to real $\overline{P}_A$ as the sample size increases. The median value showed similar behavior to the uniform distribution case, predicting nearly constant values. The ABOI method showed again a tendency to be close to real $\overline{P}_A$, but with a higher variability than for uniform distribution.

Figure 7 shows that our proposed ABOI method obtained consistently lower RMSE values than the other methods, with a remarkable improvement over arithmetic mean. Therefore, these experiments in a simulated scenario showed that the ABOI method is more reliable and more independent of the spatial distribution of samples at estimating the mathematical expectation of signal strength.

These results are also consistent with the mathematical foundations presented in Section 5, as ABOI performed well at estimating $\overline{P}_A$ in a nonuniform distribution scenario, which was close to the spatial distribution of crowdsourced measurements. In addition, as expected due to Proposition 2.2, spatial distribution based on social network theory did not satisfy the conditions required by the arithmetic mean to properly estimate $\overline{P}_A$.
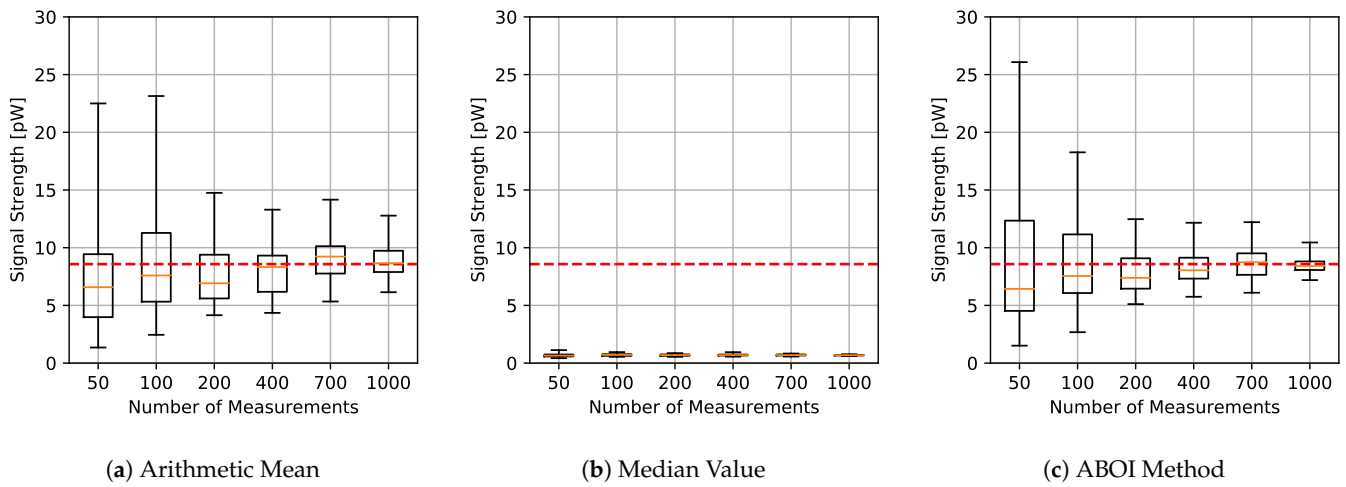
(**a**) Arithmetic Mean

(**b**) Median Value

(**c**) ABOI Method

Figure 4. Simulated scenario. Boxplots for $\overline{P}_A$ prediction using the three aggregation methods and different sample sizes, selected by uniform distribution. Real $\overline{P}_A$ value in red line.
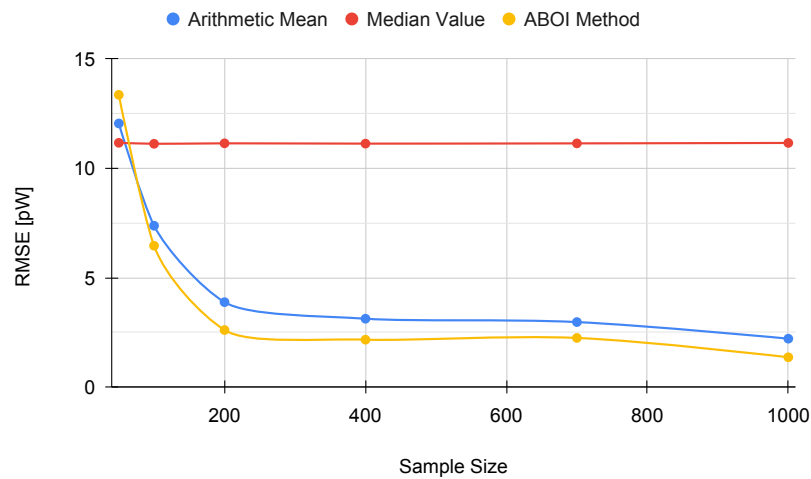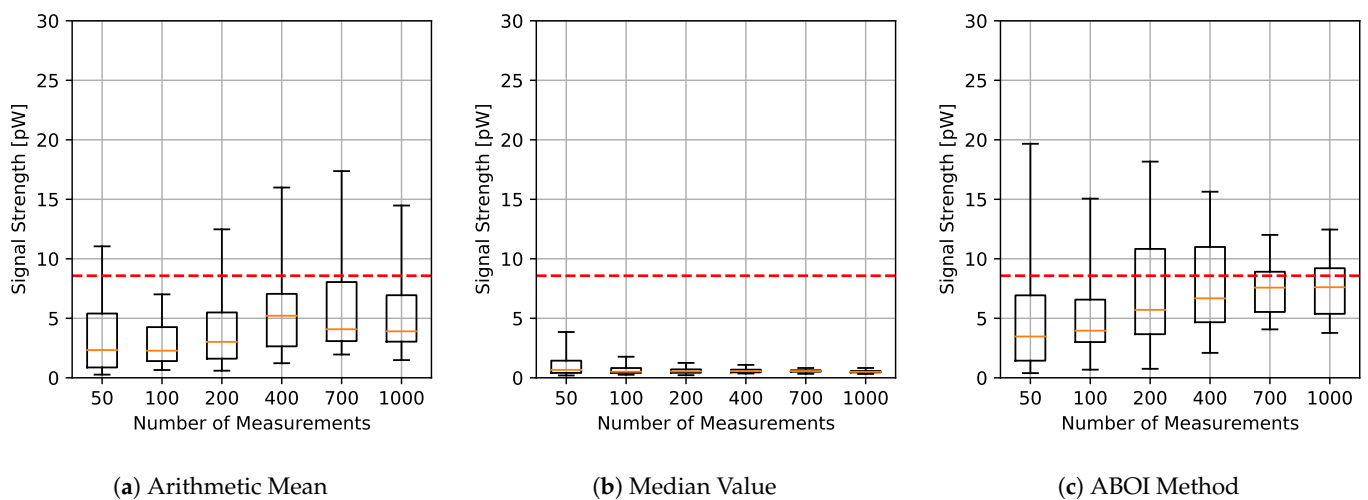

Figure 5. Simulated scenario. RMSE for $\overline{P}_A$ prediction for different sample sizes with uniform distribution.

(**a**) Arithmetic Mean

(**b**) Median Value

(**c**) ABOI Method

Figure 6. Simulated scenario. Boxplots for $\overline{P}_A$ prediction using the three aggregation methods and different sample sizes, selected by distribution based on social network theory. Real $\overline{P}_A$ value in red line.
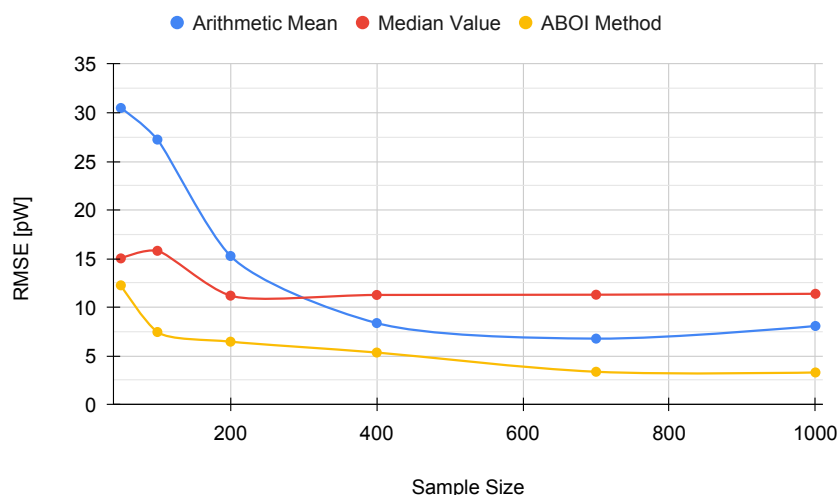
Figure 7. Simulated scenario. RMSE for $\overline{P}_A$ prediction for different sample sizes with distribution based on social network theory.

6.2. Real Data

To test the aggregation methods using real data, we developed a very minimalist Android application to take signal-strength measurements with a densely time interval. The application was designed to run every 0.5 s. During each execution, the application used Android's Telephony Manager to access information about the current cell being used by the device for network signaling. Thus, the Telephony Manager provided a CellIdentity object to obtain cell identifiers and a CellSignalStrength object to obtain the technology-specific signal strength in dBm. Along with this cell-related information, the application also stored the current location (latitude and longitude) with the highest accuracy possible.

During a period of 2 consecutive hours, we took nearly 24,000 signal-strength measurements around the vicinity of a single LTE BTS (*eNodeB*) located in a residential area, using two different mobile devices. The received signal-strength measurements densely covered an area of 140 m × 170 m near the BTS, as shown in Figure 8a. To calculate $\overline{P}_A$ value, we aligned the real measurements into a fine-grained grid G with 1 m spacing, obtaining the spatial field shown in Figure 8b. Then, we calculated the ground truth $\overline{P}_A$ as a Riemann sum of all values in G.
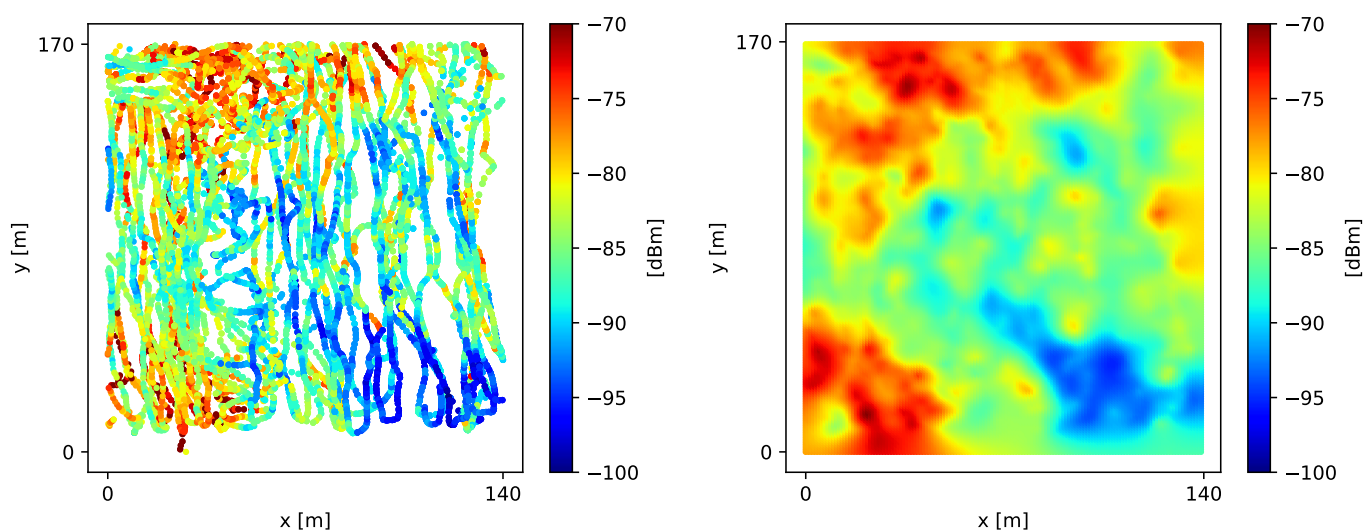
(**a**) Position of real measurements (**b**) Interpolated spatial field

Figure 8. Real signal strength around the vicinity of a single LTE BTS. Color represents the dBm value of samples.

As with the simulation case, we performed experiments for different sample sets of sizes 25, 50, 100, 200, 350 and 500. We also considered both spatial distribution methodologies: uniform distribution and based on social network theory. As for the simulation experiments, we repeated each experiment 40 times.

The results for experiments using uniform distribution are shown in Figure 9. We found that the behavior of the three methods was similar to the behavior shown by themselves in the simulation case with uniform distribution (Figure 4). The arithmetic mean and the ABOI method presented low and similar variability and a fast convergence to the calculated value of $\overline{P}_A$, where the ABOI method obtained slightly closer estimations to $\overline{P}_A$. The median value also showed coincident behavior with simulation case, predicting nearly constant and low values far from $\overline{P}_A$.

Figure 10 confirms our analysis, as both arithmetic mean and our proposed ABOI method obtained similar RMSE values, outperforming the median value. It is important to remember that as stated in Section 4.1, uniform spatial distribution is the best case for the arithmetic mean, and therefore, its good performance was expected.
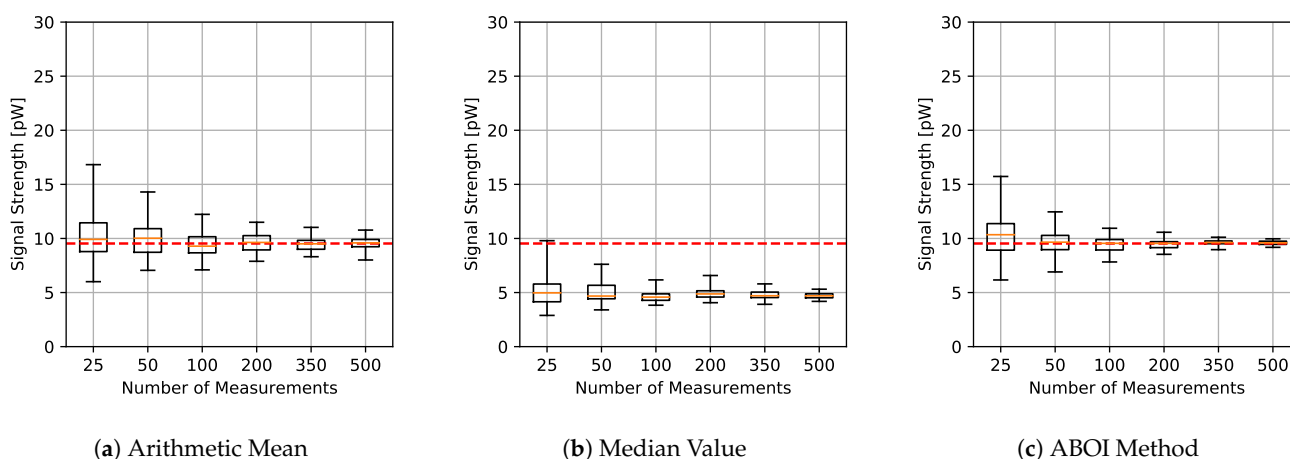

(**a**) Arithmetic Mean (**b**) Median Value (**c**) ABOI Method

Figure 9. Real scenario. Boxplots for $\overline{P}_A$ prediction using the three aggregation methods and different sample sizes, selected by uniform distribution. Calculated $\overline{P}_A$ value in red line.
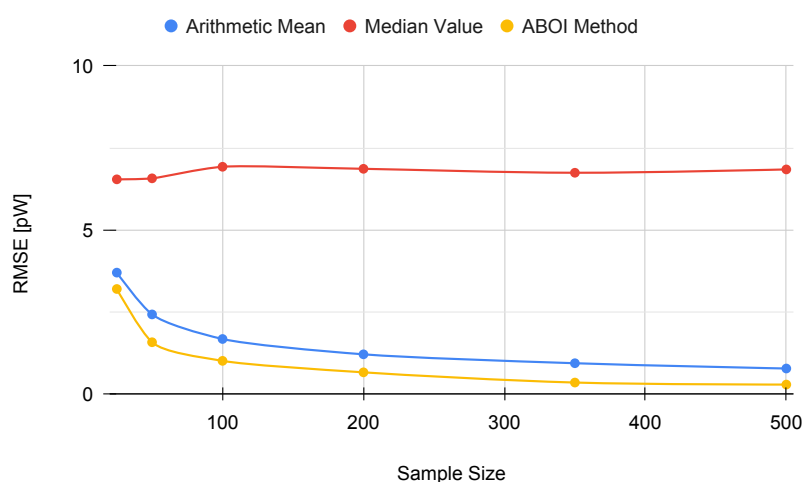

Figure 10. Real scenario. RMSE for $\overline{P}_A$ prediction for different sample sizes with uniform distribution.

As for the simulated scenario, these results are coherent with the intuition of Preposition 2.1, as in the case of uniform spatial distribution, both ABOI and the arithmetic mean performed well at estimating $\overline{P}_A$.

The results for experiments with spatial distribution based on social network theory are shown in Figure 11. The arithmetic mean showed higher variability and worse estimations of $\overline{P}_A$ in relation to the previous case. The median value tended to predict low values. Our proposed ABOI method showed a similar behavior to the uniform distribution case, showing a clear convergence to $\overline{P}_A$. It also presented lower variability than the arithmetic mean.
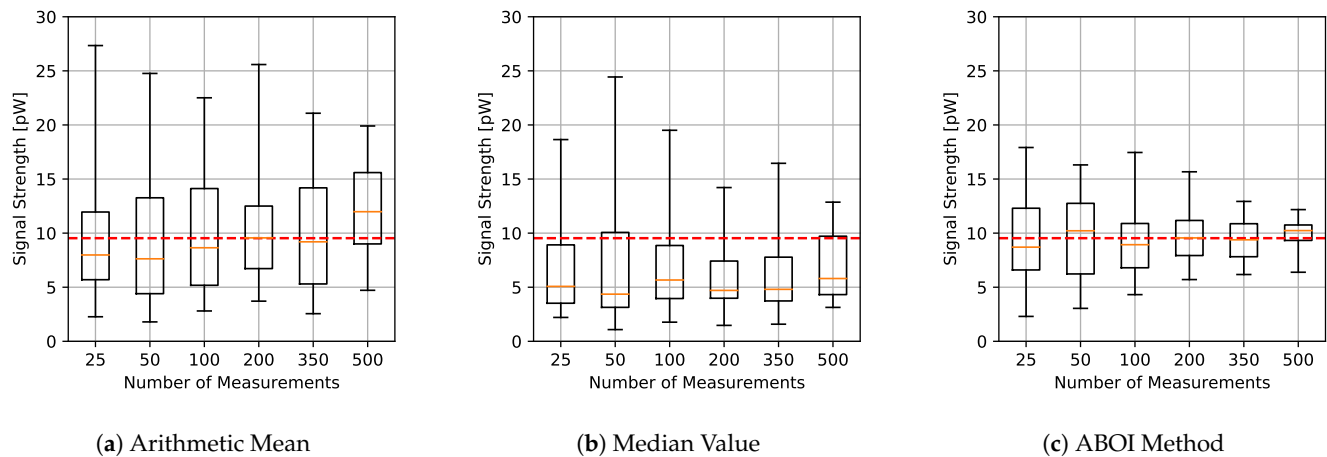

(**a**) Arithmetic Mean   (**b**) Median Value   (**c**) ABOI Method

Figure 11. Real scenario. Boxplots for $\overline{P}_A$ prediction using the three aggregation methods and different sample sizes, selected by distribution based on social network theory. Calculated $\overline{P}_A$ value in red line.

Figure 12 shows that the ABOI method obtained consistently lower RMSE values than the other methods, with a clear improvement over arithmetic mean. Unlike the arithmetic mean, our proposed method obtained more stable RMSE values at using both spatial distribution scenarios.
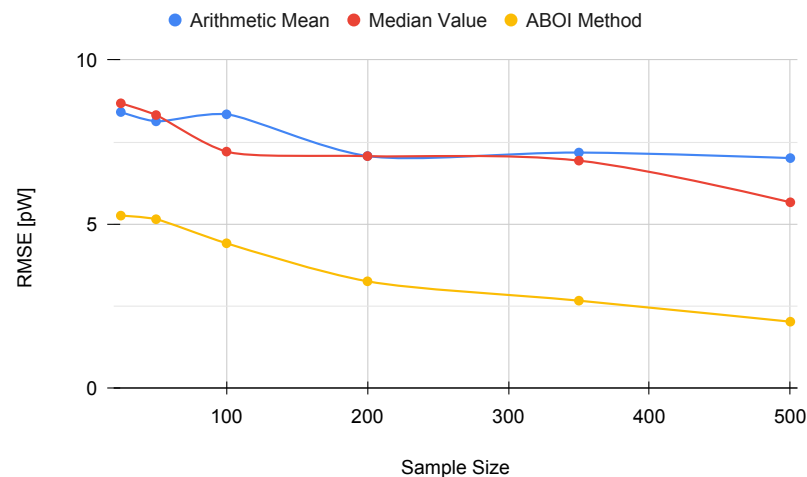

Figure 12. Real scenario. RMSE for $\overline{P}_A$ prediction for different sample sizes with distribution based on social network theory.

These results also agree with the mathematical foundations presented in Section 5 and with the obtained results in the simulated scenario, i.e., the spatial distribution based on social network theory allowed ABOI to perform well at estimating $\overline{P}_A$. In contrast, this spatial distribution did not satisfy the conditions required by the arithmetic mean to properly estimate $\overline{P}_A$.

Therefore, these experiments in a real scenario showed that our proposed method is more reliable and more independent of the spatial distribution of samples at estimating the mathematical expectation of signal strength.

Regarding the algorithms' runtime performance, we measured their execution time using a 3.4 GHz quad-core processor (Intel® Core™ i5-3570) with 12 GB RAM. For all methods, the execution time increased along with the number of measurements. On the one hand, the execution of the median value and the arithmetic mean never exceeded 0.01 s (both in the simulated case and in the real case). On the other hand, the execution of the ABOI method never exceeded 2 s. These results are an example of a well-known trade-off between estimation goodness and complexity. However, in our particular case, reducing the error at estimating $\overline{P}_A$ is much more relevant than reducing the time needed to compute the estimation, considering that the ABOI method's runtime is still very low. Therefore, we do not consider the execution time as a drawback of our method.

File S1 contains all data and materials necessary for the reproducibility of experimental results. In addition, File S2 includes all the RMSE figures, but taking into account higher numbers of measurements, to visually clarify that the ABOI method will further reduce the error in our experiments.

7. Conclusions

In this paper, we first presented the physical and mathematical formalities about how signal-strength values must be handled at applying mathematical operations in a scientific and academic environment to avoid some common sources of error. We formally showed why some simple tasks as averaging and comparing signal-strength values are usually performed in contradiction with some scientific principles due to indiscriminate use of log-scaled values, which leads to errors in the analysis of experimental data, and therefore, to making wrong conclusions.

In addition, we presented a novel method based on interpolation to aggregate signal-strength samples into one representative value to estimate the mathematical expectation of signal strength in an area. This method is shown to present solid mathematical foundations to be employed on real Mobile Crowdsourcing scenarios.

Our proposed ABOI method outperformed other commonly used aggregation methods as arithmetic mean and median value, mainly because it was shown to be more independent of some Mobile Crowdsourcing data difficulties such as nonuniform spatial distribution of the samples, the potentially low number of measurements and the inaccuracy of end-user devices. By using this method, we computed more reliable estimations of the mathematical expectation of signal strength, in both simulated and real scenarios.

We conclude that for most Mobile Crowdsourcing scenarios, our proposed ABOI method should be preferred over the other methodologies.

As this paper was our first attempt to challenge existing assumptions about signal-strength aggregation, and to propose a novel method that performs better than other used algorithms, we performed some simplifications to the problem of estimating the mathematical expectation of signal strength in our experimental scenarios. However, we did consider some important challenges found in the wild, as long-term attenuation due to path loss, medium-term variation caused by obstacles in the area, and inaccuracy of end-users hardware, reflected on GPS locations and measured signal-strength levels. All these Mobile Crowdsourcing data difficulties were present in both simulated and real experimental scenarios.

As future work several related studies that followed some of the wrong methodologies presented in this paper can be repeated by properly handling signal-strength values. Therefore, we could investigate and quantify the induced impact on the results of an incorrect mathematical treatment. In addition, the ABOI method could be improved by considering more complex simulated scenarios, as areas with multiple antennas, and taking into account small scale fading caused by multipath propagation, and short-term attenuation fluctuations due to time variance in the channel. Additionally, as mentioned

in Section [4.3](#), our proposed method could be improved by using a more complex and accurate interpolation algorithm.

Abbreviations

The following abbreviations are used in this manuscript:

QoS	Quality of Service
ABOI	Average Based on Interpolation
TCP	Transmission Control Protocol
BTS	Base Transceiver Station
ASU	Arbitrary Strength Unit
OK	Ordinary Kriging
OKD	Ordinary Kriging Detrending
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MASE	Mean Absolute Scaled Error
MSE	Mean Square Error
RMSE	Root Mean Squared Error
LTE	Long-Term Evolution

Appendix A. Proof of Mathematical Expression ([16](#))

It is clear that

$$\mathbb{E}\left[\max_{i\in[1:m]}\left|P(x_i)-I_N(x_i)\right|\right] \leq \mathbb{E}\left[\sup_{x\in A}\left|P(x)-I_N(x)\right|\right] \tag{A1}$$

Then, Corollary 1 of Wang et al. [50] holds that selecting $p = 1$,

$$\mathbb{E}\left[\sup_{x\in A}\left|P(x)-I_N(x)\right|\right] = \mathcal{O}(P_{\Phi,X}\log^{1/2}(1/P_{\Phi,X})), \text{ as } P_{\Phi,X}\to 0$$

It can be noticed that function $f(x) = x\log^{1/2}(1/x)$ satisfies the following limit:

$$\lim_{x\to 0^+} f(x) = 0$$

Therefore,

$$\lim_{P_{\Phi,X}\to 0^+}\mathbb{E}\left[\sup_{x\in A}\left|P(x)-I_N(x)\right|\right] = 0 \tag{A2}$$

Considering that power measurements $P(\cdot)$ over a rectangle area A can be modeled by Gaussian Processes (hypothesis $(\star)$ of Theorem 2), then Theorem 11.22 of Wendland [51] states that *exist positive constants c and h_0 depending only on A such that $P_{\Phi,X} \leq h_N^{c/h_N}$ provided that $h_N \leq h_0$.*

It can be seen that function $f(x) = x^{c/x}$ satisfies the following limit:

$$\lim_{x \to 0^+} f(x) = 0$$

Consequently, it can be verified that

$$\lim_{h_N \to 0^+} h_N^{c/h_N} = 0$$

and taking into account that Theorem 11.22 of Wendland [51] states that $P_{\Phi,X} \leq h_N^{c/h_N}$, we have

$$h_N \to 0^+ \implies P_{\Phi,X} \to 0^+$$

Then, Equation (A2) can be rewritten in terms of h_N as

$$\lim_{h_N \to 0^+} \mathbb{E}\left[\sup_{x \in A} |P(x) - I_N(x)| \right] = 0 \tag{A3}$$

Finally, by plugging (A3) into (A1) we obtain the desired expression in Equation (16):

$$\lim_{h_N \to 0} \mathbb{E}\left[\max_{i \in [1:m]} |P(x_i) - I_N(x_i)| \right] = 0$$

References

1. Sonntag, S.; Schulte, L.; Manner, J. Mobile network measurements-It's not all about signal strength. In Proceedings of the 2013 IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, China, 7–10 April 2013; pp. 4624–4629.
2. Huang, J.; Xu, Q.; Tiwana, B.; Mao, Z.M.; Zhang, M.; Bahl, P. Anatomizing application performance differences on smartphones. In Proceedings of the 8th international conference on Mobile systems, applications, and services, San Francisco, CA, USA, 15–18 June 2010; pp. 165–178.
3. Nikravesh, A.; Choffnes, D.R.; Katz-Bassett, E.; Mao, Z.M.; Welsh, M. Mobile network performance from user devices: A longitudinal, multidimensional analysis. In Proceedings of the International Conference on Passive and Active Network Measurement, Los Angeles, CA, USA, 10–11 March 2014; pp. 12–22.
4. Schulman, A.; Navda, V.; Ramjee, R.; Spring, N.; Deshpande, P.; Grunewald, C.; Jain, K.; Padmanabhan, V.N. Bartendr: A practical approach to energy-aware cellular data scheduling. In Proceedings of the sixteenth annual international conference on Mobile computing and networking, Chicago, IL, USA, 20–24 September 2010; pp. 85–96.
5. Oseni, O.F.; Popoola, S.I.; Abolade, R.O.; Adegbola, O.A. Comparative analysis of received signal strength prediction models for radio network planning of GSM 900 MHz in Ilorin, Nigeria. *Int. J. Innov. Technol. Explor. Eng.* **2014**, *4*, 45–50.
6. Popoola, S.I.; Atayero, A.A.; Faruk, N. Received signal strength and local terrain profile data for radio network planning and optimization at GSM frequency bands. *Data Brief* **2018**, *16*, 972–981. [CrossRef] [PubMed]
7. 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, Japan, 13–17 May 2013; pp. 35–40.
8. Sung, K.; Biswas, J.; Learned-Miller, E.; Levine, B.N.; Liberatore, M. Server-Side Traffic Analysis Reveals Mobile Location Information over the Internet. *IEEE Trans. Mob. Comput.* **2018**, *18*, 1407–1418. [CrossRef]
9. Wormald, C. Predicting User Availability from Aggregated Signal Strength Data. U.S. Patent 8,396,470, 12 March 2013.
10. Madariaga, D.; Panza, M.; Bustos-Jimenéz, J. I'm Only Unhappy when it Rains: Forecasting Mobile QoS with Weather Conditions. In Proceedings of the 2018 Network Traffic Measurement and Analysis Conference (TMA), Vienna, Austria, 26–29 June 2018; pp. 1–6.
11. Sabu, S.; Renimol, S.; Abhiram, D.; Premlet, B. Effect of rainfall on cellular signal strength: A study on the variation of RSSI at user end of smartphone during rainfall. In Proceedings of the 2017 IEEE Region 10 Symposium (TENSYMP), Cochin, India, 14–16 July 2017; pp. 1–4.
12. Rosen, S.; Lee, S.j.; Lee, J.; Congdon, P.; Mao, Z.M.; Burden, K. MCNet: Crowdsourcing wireless performance measurements through the eyes of mobile devices. *IEEE Commun. Mag.* **2014**, *52*, 86–91. [CrossRef]

13. Marina, M.K.; Radu, V.; Balampekos, K. Impact of indoor-outdoor context on crowdsourcing based mobile coverage analysis. In Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges, London, UK, 17–21 August 2015; pp. 45–50.

14. Lin, X.Y.; Ho, T.W.; Fang, C.C.; Yen, Z.S.; Yang, B.J.; Lai, F. A mobile indoor positioning system based on iBeacon technology. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 4970–4973.

15. Mojisola, D.F.; Gbolahan, K. Participatory Analysis of Cellular Network Quality of Service. *Int. J. Comput. ICT Res.* **2015**, *9*, 25–40

16. Valenzuela, R.A.; Landron, O.; Jacobs, D. Estimating local mean signal strength of indoor multipath propagation. *IEEE Trans. Veh. Technol.* **1997**, *46*, 203–212. [CrossRef]

17. Wong, D.; Cox, D.C. Estimating local mean signal power level in a Rayleigh fading environment. *IEEE Trans. Veh. Technol.* **1999**, *48*, 956–959. [CrossRef]

18. Lee, W.C. Estimate of local average power of a mobile radio signal. *IEEE Trans. Veh. Technol.* **1985**, *34*, 22–27. [CrossRef]

19. Musolesi, M.; Mascolo, C. Designing mobility models based on social network theory. *ACM Sigmobile Mob. Comput. Commun. Rev.* **2007**, *11*, 59–70. [CrossRef]

20. Li, Z.; Nika, A.; Zhang, X.; Zhu, Y.; Yao, Y.; Zhao, B.Y.; Zheng, H. Identifying value in crowdsourced wireless signal measurements. In Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, Perth, Australia, 3–7 April 2017; pp. 607–616.

21. Ng, J.K.Y.; Lam, K.Y.; Cheng, Q.J.; Shum, K.C.Y. An effective signal strength-based wireless location estimation system for tracking indoor mobile users. *J. Comput. Syst. Sci.* **2013**, *79*, 1005–1016. [CrossRef]

22. Bahl, P.; Padmanabhan, V.N. RADAR: An in-building RF-based user location and tracking system. In Proceedings of the Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No. 00CH37064), Tel Aviv, Israel, 26–30 March 2000; pp. 775–784.

23. Faggiani, A.; Gregori, E.; Lenzini, L.; Luconi, V.; Vecchio, A. Smartphone-based crowdsourcing for network monitoring: Opportunities, challenges, and a case study. *IEEE Commun. Mag.* **2014**, *52*, 106–113. [CrossRef]

24. Mankowitz, J.D.; Paverd, A.J. Mobile device-based cellular network coverage analysis using crowd sourcing. In Proceedings of the 2011 IEEE EUROCON-International Conference on Computer as a Tool, Lisbon, Portugal, 27–29 April 2011; pp. 1–6.

25. Talvitie, J.; Lohan, E.S. Modeling received signal strength measurements for cellular network based positioning. In Proceedings of the 2013 International Conference on Localization and GNSS (ICL-GNSS), Turin, Italy, 25–27 June 2013; pp. 1–6.

26. Molinari, M.; Fida, M.R.; Marina, M.K.; Pescape, A. Spatial interpolation based cellular coverage prediction with crowdsourced measurements. In Proceedings of the 2015 ACM SIGCOMM Workshop on Crowdsourcing and Crowdsharing of Big (Internet) Data, London, UK, 17–21 August 2015; pp. 33–38.

27. Fida, M.R.; Lutu, A.; Marina, M.K.; Alay, Ö. Zipweave: Towards efficient and reliable measurement based mobile coverage maps. In Proceedings of the IEEE INFOCOM 2017-IEEE Conference on Computer Communications, Atlanta, GA, USA, 1–4 May 2017; pp. 1–9.

28. Santos, I.; Murillo-Fuentes, J.J.; Djurić, P.M. Recursive Estimation of Dynamic RSS Fields Based on Crowdsourcing and Gaussian Processes. *IEEE Trans. Signal Process.* **2019**, *67*, 1152–1162. [CrossRef]

29. Lalanne, F.; Aguilera, N.; Graves, A.; Bustos, J. Adkintun Mobile: Towards using personal and device context in assessing mobile QoS. In Proceedings of the 2015 International Wireless Communications and Mobile Computing Conference (IWCMC), Dubrovnik, Croatia, 24–28 August 2015; pp. 49–54.

30. Ricker, G.; Williams, J. Averaging logarithms for detection and estimation (Corresp.). *IEEE Trans. Inf. Theory* **1974**, *20*, 378–382. [CrossRef]

31. Sonin, A.A. *The Physical Basis of Dimensional Analysis*; Department of Mechanical Engineering, MIT: Cambridge, MA, USA, 2001; pp. 1–57.

32. Bridgman, P.W. The use of dimensional formulas in changing units. In *Dimensional Analysis*; Yale University Press: London, UK, 1922; Chapter 3.

33. Bridgman, P.W., Dimensional formulas. In *Dimensional Analysis*; Yale University Press: London, UK, 1922; Chapter 2.

34. Mitchell, S.K. Comment on "Linear versus logarithmic averaging". *J. Acoust. Soc. Am.* **1967**, *41*, 863–864. [CrossRef]

35. Hershey, R.L. Analysis of the difference between log mean and mean log averaging. *J. Acoust. Soc. Am.* **1972**, *51*, 1194–1197. [CrossRef]

36. Han, Z.; Liao, J.; Qi, Q.; Sun, H.; Wang, J. Radio Environment Map Construction by Kriging Algorithm Based on Mobile Crowd Sensing. *Wirel. Commun. Mob. Comput.* **2019**, *2019*. [CrossRef]

37. Casas, P.; Seufert, M.; Wamser, F.; Gardlo, B.; Sackl, A.; Schatz, R. Next to you: Monitoring quality of experience in cellular networks from the end-devices. *IEEE Trans. Netw. Serv. Manag.* **2016**, *13*, 181–196. [CrossRef]

38. Cheng, C.C.; Hsiu, P.C. Extend your journey: Considering signal strength and fluctuation in location-based applications. *IEEE/ACM Trans. Netw.* **2014**, *23*, 451–464. [CrossRef]

39. Luo, C.; Cheng, L.; Chan, M.C.; Gu, Y.; Li, J.; Ming, Z. Pallas: Self-bootstrapping fine-grained passive indoor localization using WiFi monitors. *IEEE Trans. Mob. Comput.* **2016**, *16*, 466–481. [CrossRef]

40. Shin, H.; Chon, Y.; Kim, Y.; Cha, H. MRI: Model-based radio interpolation for indoor war-walking. *IEEE Trans. Mob. Comput.* **2014**, *14*, 1231–1244. [CrossRef]

41. Kolyaie, S.; Yaghooti, M.; Majidi, G. Analysis and simulation of wireless signal propagation applying geostatistical interpolation techniques. *Arch. Fotogram. Kartogr. Teledetekcji* **2011**, *22*, 261–270.

42. Frömmgen, A.; Heuschkel, J.; Jahnke, P.; Cuozzo, F.; Schweizer, I.; Eugster, P.; Mühlhäuser, M.; Buchmann, A. Crowdsourcing measurements of mobile network performance and mobility during a large scale event. In Proceedings of the International Conference on Passive and Active Network Measurement, Heraklion, Greece, 31 March–1 April 2016; pp. 70–82.

43. Konak, A. A kriging approach to predicting coverage in wireless networks. *Int. J. Mob. Netw. Des. Innov.* **2009**, *3*, 65–71. [CrossRef]

44. Murray, S. Beware of spectrum analyzer power averaging techniques. *Microwaves* **2006**, *45*, 57–66.

45. Alimpertis, E.; Markopoulou, A.; Butts, C.; Psounis, K. City-Wide Signal Strength Maps: Prediction with Random Forests. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 2536–2542.

46. Henry, G.T. *Practical Sampling*; Sage Publications: London, UK, 1990; Volume 21.

47. Ibrahim, M.; Parsons, J. Signal strength prediction in built-up areas. Part 1: Median signal strength. *IEE Proc. Commun. Radar Signal Process.* **1983**, *130*, 377–384. [CrossRef]

48. Parsons, J.; Ibrahim, M.; Samuel, R. Median signal strength prediction for mobile radio propagation in London. *Electron. Lett.* **1980**, *16*, 172–173. [CrossRef]

49. Janaswamy, R. Median Signal Predictions. In *Radiowave Propagation and Smart Antennas for Wireless Communications*; Springer: Boston, MA, USA, 2002; Chapter 3; pp. 67–97.

50. Wang, W.; Tuo, R.; Jeff Wu, C. On prediction properties of kriging: Uniform error bounds and robustness. *J. Am. Stat. Assoc.* **2020**, *115*, 920–930. [CrossRef]

51. Wendland, H. *Scattered Data Approximation*; Cambridge University Press: Cambridge, UK, 2004; Volume 17.

52. Niederreiter, H. *Random Number Generation and Quasi-Monte Carlo Methods*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1992; Volume 63.

53. Pronzato, L. Minimax and maximin space-filling designs: Some properties and methods for construction. *J. Société Française Stat.* **2017**, *158*, 7–13.

54. Li, X.; Mähönen, P. Grid based cooperative spectrum sensing in cognitive networks under correlated shadowing. In Proceedings of the 2012 7th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), Stockholm, Sweden, 18–20 June 2012; pp. 350–355.

55. Portelinha, J.; Martins, F.; Cardieri, P. Effects of correlated shadowing on cooperative spectrum sensing. In Proceedings of the International Workshop on Telecommunications (IWT), Santa Rita do Sapucaí, Brazil, 6–9 May 2013.

56. Hähnel, B.F.D.; Fox, D. Gaussian processes for signal strength-based location estimation. In Proceeding of Robotics: Science and Systems, Philadelphia, PA, USA, 16–19 August 2006.

57. Romero, D.; Kim, S.J.; Giannakis, G.B.; López-Valcarce, R. Learning power spectrum maps from quantized power measurements. *IEEE Trans. Signal Process.* **2017**, *65*, 2547–2560. [CrossRef]