Statistical Characterization of the 2.4 GHz Radio Channel for WSN in Indoor Office Environments

Dmitrii Kirov, Roberto Passerone, and Massimo Donelli

***Abstract*—In this paper we present the results and the assessment of a statistical analysis of the wireless sensor networks (WSN) radio channel in indoor office scenarios. The work is based on an extensive set of received signal strength (RSS) measurements collected within different typical indoor spaces and node placement scenarios. Results are compared with corresponding values from the log-distance model, which is the most widely used in WSN simulators. Channel temporal stability is also analyzed. The analysis reveals interesting regularities within the measured data and as expected we observe a highly non-stationary behavior of the RSS. We also propose several improvements for the channel model of WSN simulators based on our observations. They can be beneficial for running more accurate network simulations for such applications as localization.**

***Index Terms*—Channel modeling, Statistical characterization, Received signal strength, Localization, Wireless sensor networks.**

I. Introduction

CHANNEL path loss modeling plays a key role in many wireless applications and in simulation tools. Existing models provide a reasonable approximation for applications in which a simplified physical layer representation does not severely compromise the overall outcome. However, for several emerging systems these models become a bottleneck for achieving adequate results. This is a problem, in particular, for indoor environments, where random factors, such as multipath propagation, may affect the signal path loss [1].

As an example, wireless sensor networks (WSNs) actively use received signal strength (RSS) measurements as a channel performance indicator because it is an easily accessible metric for off-the-shelf sensor nodes. In particular, considerable part of localization applications for WSN are based on the RSS. An accurate knowledge of the underlying physical channel plays a paramount role in these systems design, calibration and deployment. Typical WSN transceivers are able to measure the RSS with a granularity of 1 [dBm]. At the same time, a deviation of 1 [dBm] might result in a distance error of $\pm$1-1.5 [m]. Real measurement uncertainty (especially in indoor environments) is typically higher because of various effects affecting the signal propagation. This could result in significantly inaccurate localization results, therefore, complex and time-consuming system prototyping and calibration are required. The latter are not always affordable. Therefore, to estimate how the system would behave in a real environment many designers rely on simulation-based design space exploration [2]. Simulation models might not capture all possible aspects of a real scenario, however, they can still be very useful in obtaining best- or worst-case boundaries [3]. Even this result is often significant for making certain design decisions (e.g., there is no need for adding extra error correction heuristics or data filtering for localization, if the model demonstrates that even in the worst case the system satisfies the requirements). The soundness of simulation results for localization applications highly depends on the underlying environment representation, and in particular on the channel model. Hence, a realistic radio channel model within a properly configured tool is of great importance for WSN designers.

The purpose of this work is twofold. First, we perform a statistical study of the 2.4 GHz WSN channel in indoor office environments in order to see if the log-distance model [4], which is currently used in most WSN simulators, provides accurate results. Second, we aim to detect regularities and/or anomalies within the collected data, if any.

Our study is based on a huge measurement campaign of collecting the RSS data in different realistic indoor spaces and node placement scenarios. We have selected several indoor spaces, which are typical for office environments (i.e., corridors, halls and office rooms). Such choice is motivated by the fact that WSNs are very frequently deployed in these environments for various purposes, from monitoring to localization [5], [6]. Also, these spaces are more accessible for experimentation with respect to industrial environments. Based on the experimental results we run the statistical characterization of the 2.4 GHz WSN channel. In particular, we compare the empirical data and model parameters derived from it with analytical values, which simulators typically provide. Also, we analyze the channel temporal stability. Finally, we study the distributions of RSS deviations and the distributions of several model parameters using the maximum likelihood method.

Our contribution includes the proposed WSN channel model improvements, which are based on our observations. The first is related to using different path loss exponent values for different mutual node placements. As a second one we propose to model the random factors of signal propagation more carefully by means of a numerical model. Also, although in this work we consider only accessible office environments without mobility, the proposed analysis methodology and model improvements could be applicable to other environments, such as industrial, characterized by high noise level

D. Kirov, R. Passerone and M. Donelli are with the Department of Information Engineering and Computer Science, University of Trento, via Sommarive 9, Trento, TN, 38123, Italy.

978-1-5090-1314-2/16/$31.00

and presence of various mechanical obstacles. This would require additional measurements, but would highly contribute to achieving more accurate results in using a WSN simulator during the design. Overall, this study can be beneficial for improving the channel models (and using them to simulate the designed system in different environments and conditions) as well as for configuring a particular system within a particular environment.

The remainder of this paper is organized as follows. In the next section we briefly review the channel models commonly used in WSN simulators as well as existing statistical studies of wireless channels. In Section III we describe our measurement methodology and experimental setup. Section IV provides the flow of statistical processing of the measured data and the results. Finally, in Section V we propose several improvements for the WSN channel model and conclude.

II. Background and related work

A. Techniques for channel modeling in WSN simulators

Several wireless channel models with different levels of accuracy, complexity and flexibility are used in wireless systems simulators. The *free space model* [4] assumes ideal propagation conditions with clear line-of-sight path between transmitter (TX) and receiver (RX) antennas and also the far-field conditions. An improvement to the free space model for long distances is the *two-ray ground reflection model*. In addition to the direct path, it also considers the signal reflected by the ground [4]. The most used channel model in WSN simulators is the *log-distance path loss model* [4]. It accounts both for propagation path loss, logarithmically decreasing, and for shadow fading effects using a probabilistic model:

$$PL(d) = PL(d_0) + 10\eta \log(d/d_0) + X_\sigma \qquad (1)$$

where d is the distance between TX and RX antennas, $PL(d_0)$ is the path loss at a reference distance d_0, which is typically one meter for indoor WSNs [7], η is the path loss exponent (PLE), which is determined either empirically or from the literature, and X_σ is a zero-mean Gaussian random variable with standard deviation σ. The Gaussian random variable accounts for shadow fading by adding some random error to the propagation path loss. The mean value of the underlying distribution is typically zero, while its standard deviation is constant throughout the simulation process. One of our hypotheses is that having a zero-mean and constant σ could be a weak assumption when different distances between TX and RX are considered in the simulation, as the behavior of the RSS error is non-stationary. We aim to demonstrate this with our experiments.

Real RSS data typically looks very random with respect to the corresponding log-distance curve. Its values can be significantly higher or lower at different distances. An example from one of our experiments is provided on Figure 1. At the same time the log-distance model and curve are claimed to provide a good approximation for analyzing the real data. Indeed, the curve is very close to a regression of the RSS data with a polynomial. Hence, the log-distance curve can be used

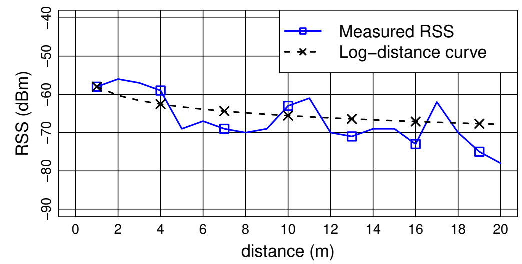

Fig. 1. Example of measured RSS data vs a corresponding log-distance curve for Corridor2, “Combined” scenario. Parameters for the curve are estimated from the measured data.

for analyzing the decay trend of the RSS with the increase of the distance. In simulations, to bring the values closer to the real data, the variable X_σ is used to account for random effects of the channel. Of course, in order to get a good approximation, the parameters of the curve ($PL(d_0)$ and σ) should be estimated from real data. Also, Formula 1 (without the random component) is typically used in localization applications to calculate the TX-RX distance from the measured RSS and parameters $PL(d_0)$ and σ. $PL(d_0)$ is usually estimated on site during system calibration, while the PLE (σ) is often not estimated but taken from best practices.

The log-distance path loss model is implemented and used in most WSN simulators, such as PASES [8], Castalia [9], MiXiM [10] and WSNet [11]. The last three also support the free space model, while the two-ray ground reflection model is supported only by WSNet.

Another powerful approach is *ray tracing* [12], which follows the signal propagation path in space, considering the physical optics rules. The main drawback of ray tracing is the required computational burden. The only tool which incorporates a (rather simplistic) ray tracing mechanism for WSN is Cooja [13].

Several approaches for considering small-scale fading are present. The MiXiM simulator was designed with high focus on mobility, therefore, small-scale fading effects due to movement are considered. The model is based on the so-called “Jakes-likes method” [10]. The Castalia simulator accounts for small-scale fading due to temporal variation of the signal. It uses numerical models based on probability density functions [14]. These models were obtained using an extensive set of RSS measurements for a body area network (BAN) [15].

B. Statistical characterization of wireless channel

Various work has been done in investigating the statistical properties of different wireless channels and environments based on empirical measurements [16]–[18]. According to them, the shadowing in the channel can be modeled as log-normal, however, the parameters of the model, namely η and σ, are random variables themselves and they have an underlying distribution. Their values could differ considerably from one case to another. Several works focus on the performance of 2.4 GHz radio links for WSN in industrial environments [19],

[20]. Such metrics as RSS, link quality indicator (LQI) and packet reception rate (PRR) are used in these works for studying the channel properties. A correlation between them is also investigated. At the same time many studies of WSN and BAN channels are focused on the RSS only [21]–[25]. In particular, results obtained by Pivato et al. suggest a non-stationary behavior of the RSS error at different distances between TX and RX [21]. Cotton and Scanlon determine best-fit distributions of RSS deviations for several line-of-sight (LOS) and no-line-of-sight (NLOS) scenarios, such as log-normal, Rayleigh and Nakagami-m [22], [23]. Smith et al. present a rigorous statistical characterization of the dynamic BAN channel based on signal strength [24], [25]. The analysis is performed over an extensive set of measurements for various scenarios, frequencies and bandwidths. Their results also confirm that different scenarios are best described with different distributions of the RSS (mostly, the Gamma distribution) with considerably different parameters.

Overall, results reported in the literature agree that the standard normal distribution is not the best description for random signal propagation effects. Therefore, the log-normal shadowing used in most WSN simulators might be an over-simplification. In many scenarios, network developers run simulations to investigate higher level functionality, such as routing, for which such model could be accurate enough. However, for such applications as localization, which uses RSS-based techniques, accurate channel modeling is of extreme importance. Naturally, absolute accuracy cannot be achieved with simulations in this case, because the measured RSS values are also sensitive to various other parameters (e.g., antenna orientation and gain, node location, type of indoor space). But improvement of the radio channel model is mandatory for accurate design and optimization tools. The analysis presented in this paper could be a basis for such improvements.

III. Experimental setup

In all of the following experiments, WSNs operating in the ISM band (2.4 GHz) were considered. The measurements were performed in several realistic indoor spaces, moreover, different node placement scenarios and antenna polarization were involved. We performed two groups of experiments:

1) *Baseline measurements*. An un-modulated carrier was considered to avoid the digital modulation, which can have a significant impact on the RSS. On the receiving side, we used a spectrum analyzer to measure the RSS value. This setup provides the baseline values for all experiments with WSN nodes.
2) *Sensor node measurements*. Here we collected the RSS using two off-the-shelf sensor nodes: transmitter and receiver. We made several experiment datasets in each indoor space with different mutual placement of nodes. For each scenario we varied the distance between the nodes with a step of 1 [m] and collected the RSS data for 5-10 minutes. With a 5 [ms] average sampling period, this resulted in roughly 120000 values in each dataset.

For the baseline measurements we used an RF synthesizer from Windfreak LLC equipped with a 2-pole antenna with a gain of 1.36 [dBi]. At the receiving side, we used a square antenna patch with the same linear polarization of the TX monopole with a gain of 3.57 [dBi], connected to a spectrum analyzer, namely Signal Hound SA124B, with a bandwidth 100 kHz - 12.5 GHz. The transmitted signal is a carrier at the frequency of 2480 MHz, which corresponds to channel 26 of the 802.15.4 standard. This channel was selected to avoid possible interference with the existing WiFi infrastructure. Such choice is common for a WSN in a building. Many existing standards implement frequency hopping and its effect on the path loss could be an interesting study as well. However, it is out of scope of this work.

WSN measurements were taken with the Z1 off-the-shelf platform produced by Zolertia with an MSP430 MCU and a CC2420 low-power radio. Two identical nodes were used for TX and for RX. Each node was enclosed in a plastic box and equipped with a 5 dBi external RP-SMA antenna. All nodes run TinyOS 2.1.2 with our testbed application. A gateway node is connected to a laptop and forwards commands to the TX/RX nodes. When the TX node receives the "start" command from the gateway, it starts sending small packets (the payload includes only a 1-byte sequence number) to the receiver. The RX node processes each packet and stores the RSS value in the log. After each experiment, logs are downloaded from the receiver via a micro-USB cable. For communication with the flash, we used the components from Trident, an open-source software for in-field connectivity assessment for WSN [26].

For all experiments, TX power was set to 0 dBm (1 mW). The period for sending packets is short (5 ms) in an attempt to observe the effect of channel temporal variation. This value is considered sufficient for observing the variation in a BAN [24]. Therefore, it should be enough for our WSN scenarios, which are much more static. Experiment duration was set to 10 minutes, but later on we reduced it by half when we observed that the channel is stable.

All measurements have been performed in the building "Polo Ferrari" of the Department of Information Engineering and Computer Science at the University of Trento, Italy, to consider realistic scenarios in the following spaces:

1) A corridor (in the following, *corridor1*), 48 x 2.8 x 2.6 [m^3], first wall - glass, second wall - gypsum plasterboard with many adjoined offices.
2) Another corridor (in the following, *corridor2*), 56 x 2.42 x 2.5 [m^3], both walls made from gypsum plasterboard, with adjoined offices on both sides.
3) A *hall*, 19.4 x 9.8 x 2.4 [m^3], side brick walls, front and back glass walls.
4) A big *office room* (11.5 x 7.5 x 3 [m^3]), one glass wall, 3 other walls - gypsum plasterboard. The room is furnished with a lot of working desks, chairs, PCs.

Our measurement scenarios aim to cover typical alternatives of WSN nodes mutual placement. We are interested in finding out regularities or anomalies in the RSS behavior across these scenarios (they are illustrated on Figure 2):

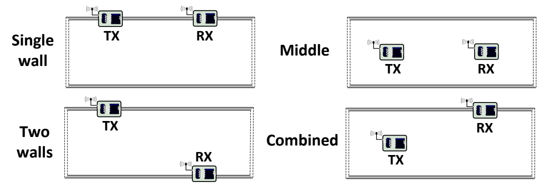

Fig. 2. WSN node mutual placement scenarios.

1) *"Single wall"*. In this scenario, both TX and RX nodes were placed on the same wall at the height of 2.25 [m] (we ran some calibration tests beforehand, placing the nodes on the bottom, middle and top of the wall and observed the strongest and most stable link at the top).
2) *"Middle"*. Both nodes were placed in the middle of the space under study at the height of 0.8 [m] (such height is very convenient for real case studies, like a bodyworn node, a PDA in a hand or a robotic device).
3) *"Two walls"*. This scenario was run only in the corridors. Nodes were located on reciprocal walls (height - 2.25 [m]). This placement scenario is typical of a WSN deployment in a corridor.
4) *"Combined"*. Was run only in the corridors. The RX node was placed on the wall (height - 2.25 [m]) and the TX node was located in the middle (height - 0.8 [m]).

Baseline measurements were made in all 4 spaces for the "Middle" scenario with both antennas placed at 0.8 [m] height. We also collected measurements near walls to compare them with the "Single wall" scenario.

Several factors are common for all scenarios. We varied the distance with a 1 [m] step to obtain distinct experimental datasets. This step has been chosen considering a good trade-off between the position updates of an object (e.g., a human) moving in a realistic scenario and the reasonable total time for collecting the measurements. The interval that we considered in our statistical investigation was [1 m, 15 m] in most cases (in some of them it was different due to space/scenario limitations). Second, we used the RSS from the minimum distance as the $PL(d_0)$ in calculations. Third, TX and RX antennas were in co-polarization condition. Finally, no mobile nodes were present in our scenarios and the LOS condition was assumed.

IV. Statistical processing workflow and results

Our results come from analyzing 16 distinct groups of data. Four are the baseline measurements, the rest are experimental datasets obtained with sensor nodes. Each group consists of separate measurement sets related to a particular distance between TX and RX. Each of the latter has roughly 60000 or 120000 values observed over a 5 min or 10 min period, respectively. TX and RX antennas were placed with the same polarization and their gain effects are removed from the data.

A. Comparing experimental and analytical data

As part of the analysis, we compare the empirical results from different scenarios with corresponding analytical values, which could be provided by a log-distance model in a WSN simulator. The latter means that channel parameters in the tool, i.e., $PL(d_0)$ and η, are configured with limited knowledge of the real channel under study. That is, the designer selects them partly or totally relying on best practices, because doing a channel characterization for deriving them empirically could be complex and time-consuming. As we demonstrate with our measurements and analysis, these best-practice parameter values in fact can be very different from those estimated from the data. Even if $PL(d_0)$ is taken from measurements (it can be done easily), different values of the PLE result in considerable difference in the curves, which entails incorrect results provided by the channel model.

For each comparison we use two curves calculated with formula (1): analytical and empirical. For the former one we select the values of the path loss exponent η from the literature. Typical value of η for indoor free space is 2, while it can be smaller for corridors (down to 1.5) and higher for furnished rooms (up to 3) when the LOS condition is assumed [7]. In industrial environments η could be bigger (up to 5-6) [4]. In this work we select analytical η to be 1.6 for the corridors, 2 for the hall and 3 for the office room. For the empirical curve we estimate η from our measurements. For doing so we calculate the linear regression for each dataset (RSS vs distance) and use its slope as an approximation for η. The value of $PL(d_0)$ for both curves is estimated from the data. For the "Combined" and "Two walls" scenarios we used 2 [m] and 3 [m], respectively, as the reference distance d_0. For all other cases 1 [m] was used. The mean RSS value from corresponding datasets was used as $PL(d_0)$. The random component X_σ of the model was set to zero to verify later on, which distributions describe the deviations of real RSS values from the log-distance curve in a best way.

For the sake of readability of the plots we do not show the measured RSS vs distance. All our measurements follow the pattern shown on Figure 1: the log-distance curve with parameters estimated from the data represents the decay trend well. The data itself is quite random, as expected. Our primary goal instead is to explore the differences between the curve drawn from the data and the one typically provided by a channel model in a simulator within a particular scenario. Random factors will be represented by a distribution, which is studied later in the paper. However, if the random component is calculated around the wrong curve, the simulation outcome might be far from reality.

1) Baseline measurements: The aforementioned comparison for baseline measurements clearly shows the general tendency of measured data to have smaller PLE values than those typically used in simulations for corresponding indoor spaces. Some illustrations are given on Figures 3 (a,b). We also compare the baseline empirical curves from different scenarios. They have the same slope (PLE), but corridor scenarios have smaller path loss (Figure 3c). This could happen due to heavy wave-guiding effects in narrow spaces like corridors.

Next, we compare sensor node and baseline empirical curves from within same scenarios. Two effects can be seen on

Figures 3 (d,e,f). First, these lines mostly have the same slope, which confirms the consistency of the sensor node measurements in relation to the baseline. Second, in all scenarios there is a significant difference between the two curves with WSN data always located higher than the baseline. There could be several possible reasons, such as hardware uncertainty of RSS measurements and/or multipath propagation effects resulting in constructive interference. Also, effects of digital modulation could affect the results. In the baseline case, we have used a sinusoid carrier, while the spectrum of the modulated signal is wider. The RSS obtained from the sensor node radio is an integral over a particular frequency interval, which could make the absolute value considerably higher than the baseline. Nevertheless, this difference does not necessarily imply huge ranging errors. The distance is estimated relatively to the $PL(d_0)$ reference, which is different for baseline and WSN measurements. The fact that the curves have similar slope implies that the ranging uncertainty based on either baseline values or measured ones would also be similar.

2) Sensor node measurements: By comparing empirical and analytical curves for WSN a difference of 5-10 [dBm] can be observed (Figures 4 a-f). In most cases, the empirical curve is higher (i.e., the path loss is smaller). One exception is the "Two walls" scenario (Figure 4f), where measured path loss, conversely, is higher than the one predicted by the analytical model. Comparison of scenarios reveals very high similarity in empirical PLE values for "Single wall", "Combined" and "Two walls" scenarios in different spaces, i.e., the curves have similar slopes (Figures 5 a-b), while for the "Middle" scenario path loss from the office room has behavior different from other spaces (Figure 5c). The office room is furnished and also WSN nodes in this scenario were placed at a lower height compared with others. Hence, the occurrence of reflections and scattering has a significant impact on the path loss.

Despite being highly similar within the same node placement scenario, empirical values of the PLE are, nevertheless, considerably dissimilar across different scenarios. This is a very important observation because it suggests that a single value of the PLE cannot accurately describe the path loss between every TX and RX within the same space if their mutual placement is different. Currently only one PLE value can be set for the whole simulation in the tools that implement the log-distance model. This might cause significant inaccuracies.

Also, one can observe from Figures 5 (a-c) the difference in the path loss within the same placement scenario in different spaces. As the PLE values are similar, this is due to the varying $PL(d_0)$ parameter. For instance, on Figure 5a corridor2 has higher values, probably, due to the wave-guiding. On Figure 5b the path loss in corridor2 is smaller than in corridor1. This is likely related to different materials of these spaces and, therefore, different electromagnetic behavior of the signal.

B. Channel stability analysis

An important part of our channel characterization is the evaluation of its temporal stability. A typical metric for this is the channel coherence time, which can be estimated with autocorrelation [15]. The latter, however, can give false negative results when most of the values in the dataset are very close to each other. Instead, in this work use a metric called *channel variation factor* [27] proposed by Zhang et al.:

$$\upsilon = \sqrt{\frac{var(x)}{\frac{1}{M}\sum_{m=0}^{M-1}|x_m|^2}} \tag{2}$$

where x is the vector of the RSS measurement sequence of length M, and $var(x)$ is the sample variance of the vector x. The RMS value of x stands in the denominator. The variation factor describes the channel stability across a particular sample of a fixed length taken from the RSS dataset. It should be noted that $0 \leq \upsilon \leq 1$. A channel with $\upsilon \leq 0.1$ can be considered temporally stable [27]. The variation factor calculated for a particular sequence provides an instantaneous variation. However, observing the values of υ on the *whole* measurement interval can give a good picture of temporal stability of the channel.

Figure 6 illustrates some variation factor results for the time-varying period of 5000 ms (the period between measurements is 5 ms, therefore, M = 1000). The reported data shows that for all measurement sets the value of υ is typically less than 0.05 and exceeds the threshold of 0.1 in an extremely rare case. It means that within any 5 second period the temporal variation is very low (and, consequently, for smaller time-varying periods; we tried smaller ones and got similar results). This is enough to conclude that the channel that we investigate is not subject to considerable temporal variations and, therefore, it is stable in all measurement scenarios. This is an interesting and useful finding because the result is similar for different indoor spaces and scenarios. Also localization systems will be less prone to random fading effects, at least for the static case.

C. First-order statistical modeling of the RSS distribution

This part of the analysis is related to fitting our datasets to several different statistical models and comparing the results in order to determine the best fit for each group. In particular, we study the deviations of measured RSS from the values provided by the log-distance model with *empirically* derived parameters $PL(d_0)$ and η. In reality, these deviations occur due to random fading effects such as shadowing. This would allow us to verify if all deviations of the RSS within different scenarios, spaces and distances can be adequately modeled by the same distribution. This is the way of implementing the log-distance model in WSN simulators: all random effects are considered by adding a random variable X_σ, which is a standard normal distribution. The value of σ is the same for all signal evaluations within a simulation. In addition, we compare the distributions of experimental data normalized to the root mean square (RMS) value of each measurement set.

Fitting distributions is performed using the maximum likelihood (ML) method. We selected six distributions for our analysis, following the similar processing flow presented by Smith et al. for BAN [24]. They are Normal, Log-normal, Gamma, Weibull, Nakagami-m and Rayleigh. It should be noted that we convert all measurements from [dBm] to [mW]

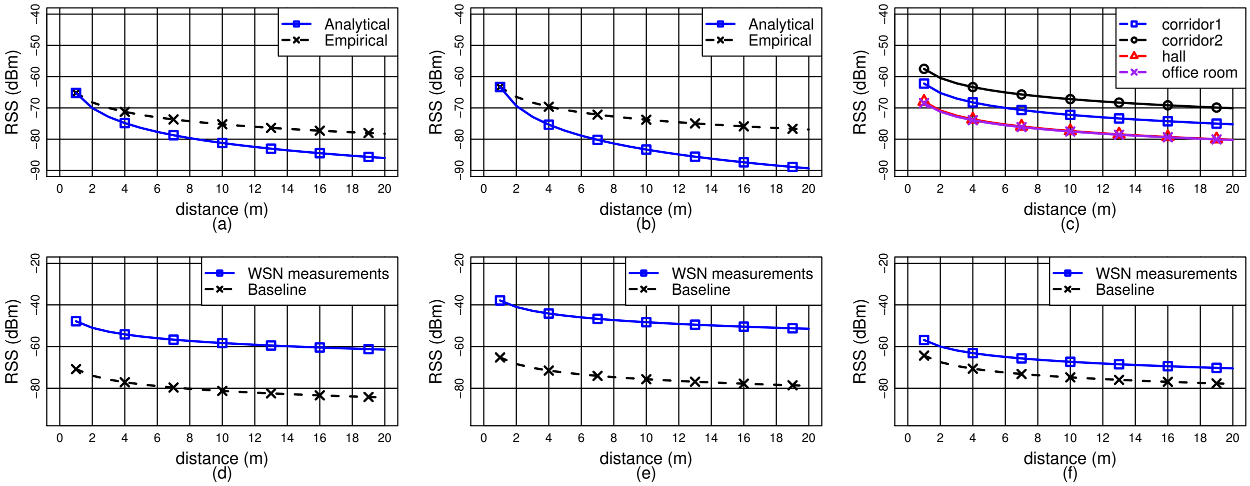

Fig. 3. Baseline measurements analysis. (a) Log-distance analytical vs empirical for Corridor1 (b) Log-distance analytical vs empirical for the Hall (c) Comparison of baseline empirical log-distance curves from studied indoor spaces (d) WSN data vs baseline (Hall, "Middle") (e) WSN data vs baseline (Corridor1, "Middle") (f) WSN data vs baseline (Corridor2, "Single wall").

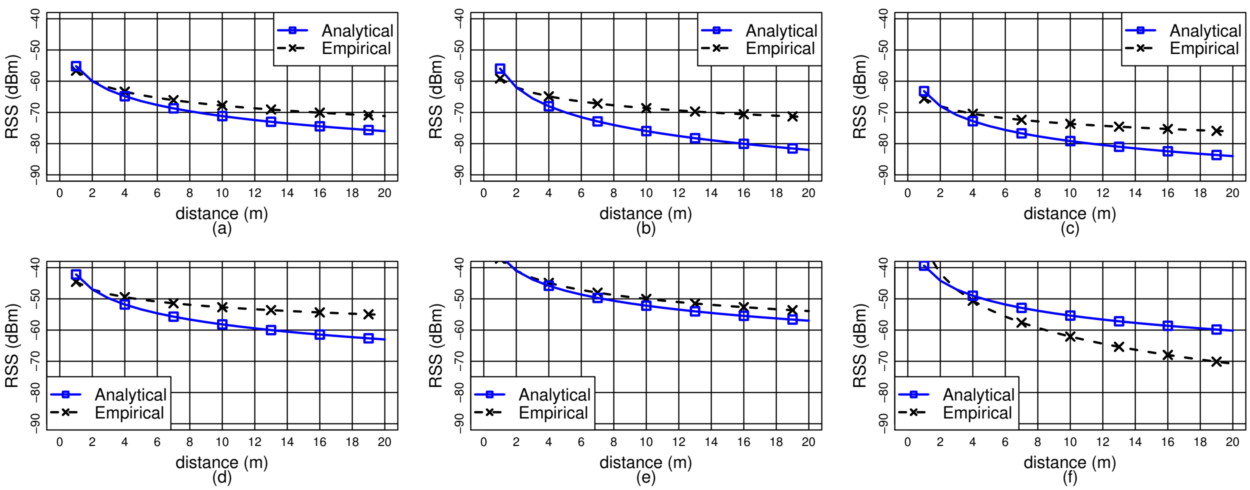

Fig. 4. Comparing analytical and empirical log-distance curves for WSN data. (a) Corridor1, "Single wall" scenario (b) Hall, "Single wall" scenario (c) Corridor1, "Combined" scenario (d) Corridor2, "Middle" scenario (e) Corridor1, "Middle" scenario (f) Corridor2, "Two walls" scenario.

before running the ML because several distributions (e.g., Gamma) require only positive values.

To compare the fitted distributions we use the Akaike information criterion [28], which is given by:

$$AIC_c = -2loglik + 2K \tag{3}$$

where $loglik$ is the value of the maximized log-likelihood function and K is the number of parameters of the corresponding distribution. In our case $K = 1$ for Rayleigh and $K = 2$ for all other distributions. This criterion allows finding a model with the minimum information loss among those that are considered. The lowest value of AIC_c is the best approximation, therefore it is of primary interest to us. In our analysis we also observed second-best fits (models with the lowest ΔAIC_c to best fits) in order to find out if several distributions are similarly good in some scenarios. We consider AIC a relevant metric for our study, because we are interested not only in accurate modeling of distribution tails (i.e., high attenuation region), but also in the values around the mean. Even small deviations of the random variable of the log-distance model might imply severe errors in applications, such as localization. At the same time, attenuation, which is below the radio sensitivity, is very unlikely for our scenarios and distances. This aspect, i.e., exploring the connected/disconnected regions, is also important, but out of scope of this work.

1) Deviations from log-distance values: To obtain these deviations, we normalize each dataset to a corresponding value from the log-distance model, i.e., subtract it from each measurement. Then we run the ML estimation. The obtained results clearly demonstrate that best fitted distributions are considerably different across scenarios and spaces. Another interesting fact is that varying the distance between TX

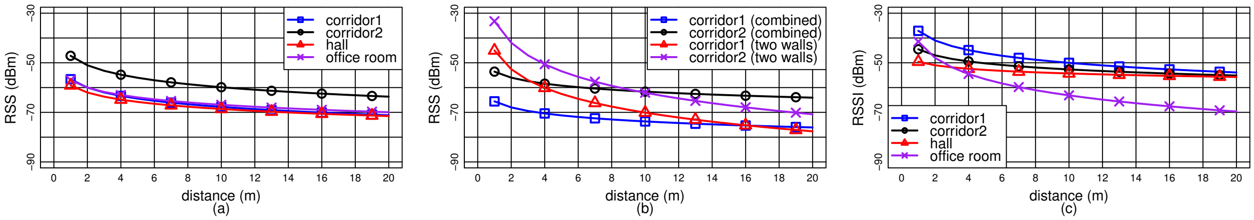

Fig. 5. Comparison of log-distance empirical curves for WSN data. (a) “Single wall” scenario (b) “Combined” and “Two walls” (c) “Middle”.

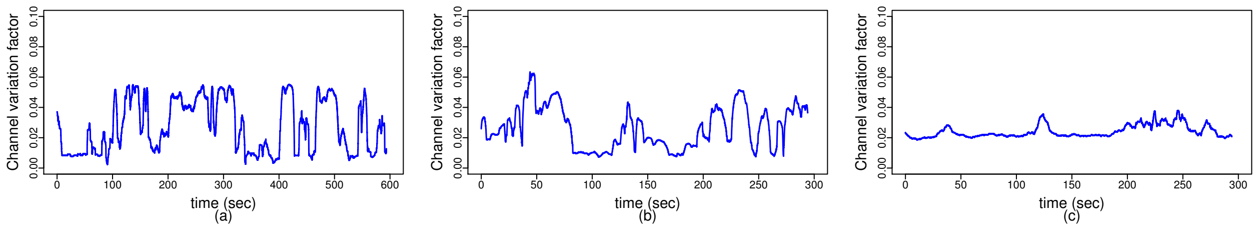

Fig. 6. Channel variation factor for a time-varying period of 5000 ms for several indoor spaces, scenarios and TX-RX distances. (a) Corridor2, ”Middle”, 12m (b) Office room, ”Middle”, 8m (c) Hall, ”Single wall”, 9m.

and RX within a certain scenario also highly affects these distributions. Even if the distribution itself is the same, its parameters differ significantly from case to case. This is a very important observation because it provides empirical evidence that a single standard normal distribution (log-normal for RSS values in [mW]; this entails that values in [dBm] have normal distribution) with fixed parameters cannot accurately describe deviations from the log-distance model. This is, however, the way of most WSN simulators, therefore, it is highly doubtful that such simulations could provide adequate results for localization applications.

In many scenarios, the Log-normal distribution is the dominating one, i.e., it provides the best fit at 50-60% of distances. In several scenarios, however, Gamma or Weibull distribution dominate. The Normal distribution has fewer occurrences, Nakagami-m is the best fit only for 3 measurement groups among all scenarios, while the Rayleigh distribution is never observed in any scenario. Due to space limitations, we do not report the numerical results for this study.

2) Deviations from the RMS: The results from previous analysis do not reveal evident regularities for making any improvement in the models. Therefore, we continued our analysis by normalizing our data by the RMS value. We then created agglomerated datasets by joining normalized values for each scenario (for example, 4 datasets for the “single wall” scenario were joined into one for all TX-RX distances) and, similarly, for each indoor space (e.g., joined the data from all placement scenarios for “corridor1”).

From the results we observe much more similarities than in the previous case. The overall percentage of Log-normal distribution is higher, but still not absolutely dominating. The following are Gamma and Normal distributions, while Weibull and Nakagami-m are observed rarely. Rayleigh is again never observed. Second-best fits could not be considered in the analysis as ΔAIC has very high values (it should be no more than 10 in order to be considered “close” to the best case [29]; our values have orders of magnitude from 10^2 to 10^4). This also states that the obtained best fits clearly outperform their closest counterparts. Comparing different agglomerates of the same placement scenario shows many similar distributions at a particular TX-RX distance, however, in many cases distribution parameters differ considerably, which does not allow the confident selection of a single distribution, representing the data. The situation is better for the data agglomerates within the same indoor space. Log-normal distribution is the most common with much more similarity in parameters across different sets. If the distributions are nevertheless different at the same TX-RX distance, we performed an additional check by means of simulation. Test datasets were generated for each distribution and their means and standard deviations were compared. This check showed that the values from these datasets are quite close to each other (within a few dBm).

Finally, taking into account the above results, we created a total agglomerate of all RMS-normalized data across all indoor spaces and scenarios and estimated the best fits. We report these results in Table I. There we can see that the Log-normal distribution often occurs, however, with different parameters, but there are also some TX-RX distances where Gamma, Normal and Nakagami-m distributions provide the best fit. These final statistical results once again support our hypothesis that a single distribution is not able to describe random RSS deviations well and, therefore, this might be a potential loss of accuracy for the log-distance path loss model.

One might argue that agglomerated results could entail substantial loss of accuracy despite the high similarity in distributions describing separate datasets. Indeed, such methods as averaging and joining data from different groups are typically error-prone. However, our selective checks, during which we

TABLE I
MLE FOR AGGLOMERATES ACROSS ALL SPACES AND SCENARIOS

Distance	Distribution	ΔAIC
1m	Normal (μ = 1.011, σ = 0.142)	7450
2m	Lognormal (μ = 0.0025, σ = 0.259)	14757
3m	Gamma (a = 27.960, b = 27.425)	5337
4m	Lognormal (μ = 0.0012, σ = 0.187)	24025
5m	Lognormal (μ = 0.0020, σ = 0.238)	53088
6m	Lognormal (μ = 0.0027, σ = 0.284)	12483
7m	Lognormal (μ = 0.0018, σ = 0.234)	80391
8m	Lognormal (μ = 0.0040, σ = 0.349)	647
9m	Lognormal (μ = 0.0056, σ = 0.416)	40594
10m	Gamma (a = 11.802, b = 11.273)	13355
11m	Gamma (a = 16.601, b = 16.070)	19764
12m	Lognormal (μ = 0.0044, σ = 0.370)	58543
13m	Lognormal (μ = 0.0019, σ = 0.243)	16629
14m	Nakagami-m (m = 7.474, ω = 0.187)	1864
15m	Lognormal (μ = 0.0019, σ = 0.255)	64069

Normal and Lognormal: μ - mean, σ - standard deviation. Gamma: a - shape, b - rate. Nakagami-m: m - shape, ω - scale.

generated values from non-agglomerated and agglomerated datasets and compared their means and standard deviations, showed that in our case results are quite similar. This means that the differences are acceptable for the radio domain. Hence, we concluded that our agglomerations are feasible.

The similarity of distributions and their parameters allowed us to agglomerate the RSS data from different scenarios at each measurement distance. However, there are notable differences in distribution parameters at different distances, which does not suggest to use a single distribution. This fact supports our claim that standard zero-mean Normal distribution does not describe random factors in the channel at different TX-RX distances equally well. The difference in the mean values of our resulting distributions can reach up to several [dBm]. By getting inaccurate results from the channel model designers might draw incorrect conclusions on the RSS uncertainty and under-design (or over-design) an RSS-based localization system in terms of accuracy.

3) Log-normal distribution parameters: We remark that around 2/3 of best-fit distributions of RSS deviations are Log-normal. Despite this fact, their parameters considerably differ. We do not observe any specific relationship between their values and scenario parameters (distance and node placement). However, they could have a distribution. It is previously reported that the parameters of the Log-normal distribution, μ and σ, are random variables themselves and have underlying models [16], [17]. It would be interesting to verify this for our scenarios. We use the ML results for all our datasets to extract the estimated μ and σ (even if Log-normal is not a best-fit). Then we create sets of these parameters for particular combinations of space and scenario (and similarly for agglomerates) and run the ML estimation over them.

Results of the estimation show that the mean (μ) parameter is distributed log-normally, while the standard deviation (σ) is Rayleigh-distributed for most (80-90%) of spaces and scenarios. Parameters of these distributions are mostly similar. This allows us to create full agglomerates of μ and σ parameters and report that the overall distribution for μ is Log-normal (with the mean of -6.078 and standard deviation of 0.444) and σ the distribution is Rayleigh (with the scale of 0.231).

V. CONCLUSIONS AND FUTURE WORK

In this paper we analyzed the 2.4 GHz wireless sensor network channel using a statistical approach based on a huge experimental measurement campaign performed in several realistic indoor scenarios. Our observations allowed us to conclude that the behavior of RSS is very non-stationary across different scenarios as expected. We prove empirically that the log-distance path loss model widely applied in existing WSN simulators does not accurately describe the path loss at different distances between TX and RX nodes. In particular, empirically derived values of the path loss exponent are considerably different from those commonly used in simulator configurations (Figures 4 a-f). Moreover, node mutual placement highly affects the PLE (Figures 5 a-c).

The WSN channel investigated in this work is not subject to considerable temporal variation, as confirmed by low values of the variation factor. Deviations of RSS values are dissimilar at different TX-RX distances. The most frequently observed statistical model, which describes these deviations in a best way, is log-normal. However, it cannot be called dominating, because its parameters may considerably vary from case to case and also other models are present (e.g., Gamma). There are notable similarities of distributions for the same TX-RX distance across different spaces and placement scenarios. This allowed us to agglomerate the normalized RSS data and summarize the results (Table I). They clearly demonstrate that using a single distribution with fixed parameters to model large- and small-scale fading effects for all scenarios may entail a potential loss of accuracy during a simulation.

On the basis of our results we propose the following improvements to the log-distance channel model in WSN simulators. First, different values of the path loss exponent can be used for different placement scenarios. During a simulation one could determine the mutual placement of nodes, for which the path loss is evaluated, and use the corresponding PLE. In particular, we noticed that the PLE for the "Two walls" scenario can be 2-3 times higher than for others. Therefore, it requires a separate PLE value to provide accurate results. Values for other placement scenarios (in this work we tried only 4 most typical) can also be configurable.

Second proposed improvement is related to modeling random factors using a certain distribution. During the evaluation of a particular path loss, one could check the distance between TX and RX and use a distribution from a corresponding distance interval instead of using the same standard normal distribution in every case. Another approach is to keep the existing X_σ random variable but allow the distribution parameters (mean μ and standard deviation σ) to be generated every time with their corresponding distributions. For example, for scenarios studied in this work μ and σ can be generated with Log-normal and Rayleigh distributions, respectively, as reported in Section IV-C3.

Currently we are working on implementing the proposed improvements in the PASES simulator [8]. In particular, instead of raising the configuration complexity by introducing different distributions and their parameters, we provide a *numerical model* for random propagation factors. We use the algorithm for temporal variation modeling from the Castalia simulator [14]. It can fetch the required value from a numerical model with a very low performance overhead. In PASES we use this approach for temporal models (same as in Castalia) and we also introduce novel *distance-dependent* numerical models for random propagation factors. The latter are generated using our numerical results from Section IV-C obtained from our measurement campaign. This approach is extensible, i.e., numerical models for other environments (e.g., industrial) can be generated from datasets, which can be obtained using our measurement and statistical characterization methodology.

Overall, our suggested improvements to the log-distance path loss model are related to enforcing more accurate values for both the deterministic part of Formula 1 ($PL(d_0)$ and σ) and the stochastic part (using proper distribution and its parameters). We believe that results and proposals from this paper will allow to have more accurate simulations of WSN applications with low effort.

Our future work is the evaluation of the improved PASES simulator. Moreover, this study is a good starting point, from which we can move our investigation to real industrial environments to draw more conclusions about the WSN radio channel and further improvements in models. Such environments are typically characterized by high electromagnetic pollution, usually at low frequencies. Also, NLOS conditions are to be studied due to the presence of strong multipath effects, which might play a dominant role in signal path loss. Furthermore, movement scenarios are to be considered in both LOS and NLOS. From the latter we expect lower temporal stability and coherence time of the channel, in particular, due to Doppler effects related to moving object or environment. This might require additional calibration of existing models.

References

[1] F. Viani, L. Lizzi, P. Rocca, M. Benedetti, M. Donelli, and A. Massa, "Object tracking through RSSI measurements in wireless sensor networks," *Electronics Letters*, vol. 44, no. 10, pp. 653–654, 2008.

[2] A. Davare, D. Densmore, L. Guo, R. Passerone, A. L. Sangiovanni-Vincentelli, A. Simalatsar, and Q. Zhu, "METROII: A design environment for cyber-physical systems," *ACM Transactions on Embedded Computing Systems*, vol. 12, no. 1s, pp. 49:1–49:31, March 2013.

[3] I. Minakov, R. Passerone, A. Rizzardi, and S. Sicari, "A comparative study of recent wireless sensor network simulators," *ACM Transactions on Sensor Networks*, 2016.

[4] T. S. Rappaport, *Wireless communications: principles and practice*. prentice hall PTR New Jersey, 1996, vol. 2.

[5] D. Brunelli, I. Minakov, R. Passerone, and M. Rossi, "POVOMON: an ad-hoc wireless sensor network for indoor environmental monitoring," in *Proceedings of the 2014 IEEE Workshop on Environmental, Energy and Structural Monitoring Systems*, Naples, Italy, September 17–18, 2014.

[6] ——, "Smart monitoring for sustainable and energy-efficient buildings: a case study," in *Proceedings of the IEEE Workshop on Environmental, Energy and Structural Monitoring Systems*, Trento, Italy, July 2015.

[7] J. D. Gibson, Ed., *The Mobile Communication Handbook*, 1st ed. CRC Press, Feb 1996.

[8] I. Minakov and R. Passerone, "PASES: An energy-aware design space exploration framework for wireless sensor networks," *Journal of Systems Architecture*, vol. 59, no. 8, pp. 626–642, Sep. 2013.

[9] A. Boulis, "Castalia: Revealing Pitfalls in Designing Distributed Algorithms in WSN," in *SenSys'07*, Sydney, Australia, November 2007.

[10] A. Köpke, M. Swigulski, D. Willkomm, and K. Wessel, "Simulating Wireless and Mobile Networks in OMNeT++ The MiXiM Vision," in *OMNeT++ Workshop '08 Marseille, France*. ICST, 2008.

[11] G. Chelius, A. Fraboulet, and E. Fleury, "Worldsens: a fast and accurate development framework for sensor network applications," in *Proceedings of the 2007 ACM symposium on Applied computing*, 2007.

[12] K. R. Schaubach, N. J. Davis IV, and T. S. Rappaport, "A ray tracing method for predicting path loss and delay spread in microcellular environments," in *Vehicular Technology Conference*, 1992, pp. 932–935.

[13] J. Eriksson, F. Österlind *et al.*, "Cooja/mspsim: Interoperability testing for wireless sensor networks," in *Proceedings of the 2Nd International Conference on Simulation Tools and Techniques*, ser. Simutools '09, ICST, Brussels, Belgium, 2009, pp. 27:1–27:7.

[14] A. Boulis, "Castalia user's manual, March 2011," http://castalia.research.nicta.com.au/index.php/en/documentation.html.

[15] D. B. Smith, A. Boulis, and Y. Tselishchev, "Efficient conditional-probability link modeling capturing temporal variations in body area networks," in *Proceedings of the 15th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems*. ACM, 2012, pp. 271–276.

[16] V. Erceg, L. J. Greenstein *et al.*, "An empirically based path loss model for wireless channels in suburban environments," *Selected Areas in Communications, IEEE Journal on*, vol. 17, no. 7, pp. 1205–1211, 1999.

[17] S. S. Ghassemzadeh, R. Jana, C. W. Rice, W. Turin, and V. Tarokh, "A statistical path loss model for in-home UWB channels," in *Ultra Wideband Systems and Technologies, Digest of Papers*, 2002.

[18] C. F. Mecklenbraüker, A. F. Molisch *et al.*, "Vehicular channel characterization and its implications for wireless system design and performance," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1189–1212, 2011.

[19] V. C. Gungor, B. Lu, and G. P. Hancke, "Opportunities and challenges of wireless sensor networks in smart grid," *Industrial Electronics, IEEE Transactions on*, vol. 57, no. 10, pp. 3557–3564, 2010.

[20] L. Tang, K.-C. Wang, Y. Huang, and F. Gu, "Channel characterization and link quality assessment of ieee 802.15. 4-compliant radio for factory environments," *Industrial Informatics, IEEE Transactions on*, vol. 3, no. 2, pp. 99–110, 2007.

[21] P. Pivato, L. Palopoli, and D. Petri, "Accuracy of RSS-based centroid localization algorithms in an indoor environment," *Instrumentation and Measurement, IEEE Transactions on*, vol. 60, no. 10, 2011.

[22] S. L. Cotton and W. G. Scanlon, "Indoor channel characterisation for a wearable antenna array at 868 MHz," in *Wireless Communications and Networking Conference*, vol. 4. IEEE, 2006, pp. 1783–1788.

[23] S. Cotton and W. Scanlon, "Characterization and modeling of the indoor radio channel at 868 MHz for a mobile bodyworn wireless personal area network," *Antennas and Wireless Propagation Letters, IEEE*, vol. 6, pp. 51–55, 2007.

[24] D. B. Smith, L. W. Hanlen, J. A. Zhang, D. Miniutti, D. Rodda, and B. Gilbert, "First-and second-order statistical characterizations of the dynamic body area propagation channel of various bandwidths," *Annals of telecommunications*, vol. 66, no. 3-4, pp. 187–203, 2011.

[25] L. Hanlen, V. Chaganti *et al.*, "Open-source testbed for body area networks: 200 sample/sec, 12 hrs continuous measurement," in *21st International Symposium on Personal, Indoor and Mobile Radio Communications Workshops*. IEEE, 2010, pp. 66–71.

[26] T. Istomin, R. Marfievici, A. L. Murphy, and G. P. Picco, "TRIDENT: In-field Connectivity Assessment for Wireless Sensor Networks," in *Proceedings of the 6th Extreme Conference on Communication and Computing (ExtremeCom)*, 2014.

[27] J. Zhang, D. B. Smith *et al.*, "Stability of narrowband dynamic body area channel," *Antennas and Wireless Propagation Letters, IEEE*, vol. 8, pp. 53–56, 2009.

[28] H. Akaike, "Information theory as an extension of the maximum likelihood principle," in *Second International Symposium on Information Theory*, 1973, p. 267.

[29] K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2002.