Combining a physical model with a nonlinear fluctuation for signal propagation modeling in WSNs

Sandy Mahfouz*, Paul Honeine*, Farah Mourad-Chehade*, Joumana Farah†, Hichem Snoussi*
*Institut Charles Delaunay (CNRS), Université de Technologie de Troyes, Troyes, France
†Department of telecommunications, Faculty of engineering, Université Saint-Esprit de Kaslik, Lebanon

***Abstract*—In this paper, we propose a semiparametric regression model that relates the received signal strength indicators (RSSIs) to the distances separating stationary sensors and moving sensors in a wireless sensor network. This model combines the well-known log-distance theoretical propagation model with a nonlinear fluctuation term, estimated within the framework of kernel-based machines. This leads to a more robust propagation model. A fully comprehensive study of the choices of parameters is provided, and a comparison to state-of-the-art models using real and simulated data is given as well.**

***Index Terms*—Distance estimation, kernel functions, multi-kernel learning, RSSI, semiparametric regression.**

I. INTRODUCTION

Received signal strength indicator (RSSI), a standard feature in most radios, has shown attractive properties that could be exploited for localization in WSNs [1]. In fact, one can take advantage of the attenuation of the signal strength with the increase of the traveled distance, to estimate the distances between sensors, using only RSSI information, without the need of additional hardware. However, estimating the exact distances using RSSI turns out to be really challenging, since the measurements of signals' powers could be significantly altered by the presence of additive noise, multipath fading, shadowing, and other interferences. Therefore, it becomes more and more important to find a mathematical model which can accurately describe the relationship between the RSSI values and the distances.

Several models have been proposed in the literature to characterize the relationship between the RSSIs and the distances. A popular one is the Okumura-Hata model, also known as the log-distance propagation model [2], [3], [4]; even though this model has many limitations, it is still widely used because of its simplicity. However, this model is basically for outdoors, since it predicts the signal strength without taking the surrounding environment into account, such as the walls and the floors. Therefore, it becomes inaccurate in cases where there is no line of sight between sensors. Different models have been proposed to overcome this problem, such as modified versions of the log-distance model [5], [6], in which the attenuations due to floors and walls are explicitly included. In other ones [7], [8], the authors determine a mathematical relationship between RSSIs and distances, without taking physical properties into consideration. To this end, they estimate an empirical model based on polynomial regression. Depending on the applications and the considered environment, several other models can also be found in the literature [9], [10], [11].

Lately, kernel machine methods have been greatly used in the context of function approximation and model fitting, since they simplify the specification of a nonparametric model, especially in the case of nonlinear data [12], [13], [14]. In addition, kernel machine methods are being widely used for semiparametric regression, to characterize relationships between data in the medical domain as in [15], or in hyperspectral imaging as in [16]. In this paper, we propose a semiparametric propagation model to define the relationship between the RSSIs and distances. The proposed model combines the well-known log-distance propagation model with a nonlinear fluctuation term, defined in a reproducing kernel Hilbert space. The nonlinear term added in the proposed model compensates for the missing term in the log-distance model, i.e., all environment's sources of noises, therefore allowing a better modeling of the RSSI/distance relationship. To our knowledge, such a model has not yet been proposed in the literature.

The rest of the paper is organized as follows. Section II gives a brief description of two well-known state-of-the-art models, along with an estimation of their parameters. Section III describes the proposed semiparametric model. Section IV provides a comparison of the proposed model to the two state-of-the-art models using real and simulated data, while Section V concludes the paper.

II. STATE-OF-THE-ART MODELS

As previously explained, many signal propagation models, that characterize the relationship between the RSSIs and the distances, are proposed in the literature. Some models take into account the physical characteristics of the environment, such as the path-loss in the well-known theoretical log-distance model [2], [3]. Others avoid the physical properties by finding a mathematical relationship based on polynomial regression [7], [8].

A training phase is necessary for both types of models in order to determine their parameters. To this end, N reference positions, denoted by $\boldsymbol{p}_\ell$, $\ell \in \{1, \ldots, N\}$, are generated uniformly or randomly in the studied environment. A stationary sensor with known position $\boldsymbol{s}$ continuously broadcasts signals in the network at a fixed initial power, and a sensor is placed consecutively at the reference positions to detect the broadcasted signals and measure their RSSIs. Let ρ_ℓ be the power received from the stationary sensor at position $\boldsymbol{s}$ by the sensor at position $\boldsymbol{p}_\ell$. Then, the distances d_ℓ between each reference position $\boldsymbol{p}_\ell$ and the stationary sensor are computed. In this way, a training set of N pairs (ρ_ℓ, d_ℓ) is obtained, that can be used for the estimation of the model's parameters.

After determining the model's parameters, one can estimate the distance between the stationary sensor at position $\boldsymbol{s}$ and a sensor at an unknown position $\boldsymbol{x}_i$ using the estimated model and the RSSI value ρ_i measured by the sensor.

A brief description of the log-distance model and the estimation of its parameters is provided in the first subsection. In the second subsection, a description of the mathematical model based on polynomial regression is provided, along with its parameters.

A. Log-distance propagation model

Both theoretical and measurements-based propagation models indicate that the received signal strength decreases logarithmically with distance. Therefore, the log-distance propagation model, that characterizes the relationship between RSSIs and distances, is expressed as follows:

$$\rho_i = \rho_0 - 10\, n_P \, \log_{10} \frac{d_i}{d_0}, \tag{1}$$

where ρ_i is the power received from the stationary sensor at position $\boldsymbol{s}$ by the sensor at position $\boldsymbol{x}_i$, ρ_0 is the power at the reference distance d_0, $d_i = \|\boldsymbol{s} - \boldsymbol{x}_i\|$ is the Euclidian distance between the position $\boldsymbol{x}_i$ of the considered node and the position $\boldsymbol{s}$ of a stationary sensor, and n_P is the path-loss exponent. The value of n_P depends on the specific propagation environment, i.e. type of construction material, architecture, location within building, dimensions of the coverage area, etc. For low values of n_P, the signal loss towards distance is low. Developing equation (1) leads to the following:

$$\log_{10} d_i = \frac{\rho_0}{10\, n_P} + \log_{10} d_0 - \frac{\rho_i}{10\, n_P}. \tag{2}$$

Now let $a = -\frac{1}{10\, n_P}$ and $b = \frac{\rho_0}{10\, n_P} + \log_{10} d_0$. Then equation (2) can be written as follows:

$$\log_{10} d_i = a\, \rho_i + b. \tag{3}$$

The parameters a and b depend on the characteristics of the environment and can be estimated from the N reference measurements using the Least-Squares (LS) method. They must be chosen in a way to minimize the sum of squared residuals, i.e. the mean squared error on the training set, given by:

$$\frac{1}{N} \sum_{\ell=1}^{N} \varepsilon_\ell^2. \tag{4}$$

Here, $\varepsilon_\ell = \log_{10} d_\ell - a\, \rho_\ell - b$, and $\ell \in \{1, \ldots, N\}$.

By taking the partial derivatives of (4) with respect to a and b and setting them to zero, one gets the following in matrix form:

$$\begin{bmatrix} \mathbf{1}^\top \boldsymbol{D} \\ \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \mathbf{1}^\top \boldsymbol{\rho} & N \\ \boldsymbol{\rho} & \mathbf{1} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}, \tag{5}$$

where $\mathbf{1}$ is the $N \times 1$ vector of ones, $\boldsymbol{D}$ is the $N \times 1$ vector whose ℓ-th entry is equal to $\log_{10} d_\ell$, and $\boldsymbol{\rho}$ is the $N \times 1$ vector whose ℓ-th entry is equal to ρ_ℓ, with $\ell \in \{1, \ldots, N\}$.

Having computed the values of a and b using (5), one can estimate the distance $\hat{d}_i$ between the considered stationary sensor at position $\boldsymbol{s}$ and a sensor at an unknown position $\boldsymbol{x}_i$ in the network using the following equation:

$$\hat{d}_i = 10^{a\, \rho_i + b},$$

where ρ_i is the power received from the stationary sensor at position $\boldsymbol{s}$ by the sensor at position $\boldsymbol{x}_i$.

B. Polynomial model

Having described the log-distance propagation model and its parameters, we introduce in this subsection the polynomial model. The objective of using a polynomial regression is to determine a mathematical relation between RSSIs and distances, without having to take physical properties into consideration. In such case, the signal propagation model is then the n-th degree polynomial given by the following:

$$d_i = a_0 + a_1\, \rho_i + a_2\, \rho_i^2 + \cdots + a_n\, \rho_i^n, \tag{6}$$

where a_j, $j \in \{0, \ldots, n\}$, are the polynomial's coefficients to be determined.

The parameters a_j should be chosen in a way to fit the n-th degree polynomial through the training set. The LS method is used here as well; therefore, we need to find the parameters a_j that allow us to minimize the mean squared error on the training set as given in (4), with $\varepsilon_\ell = d_\ell - a_0 - a_1\, \rho_\ell - a_2\, \rho_\ell^2 + \cdots - a_n\, \rho_\ell^n$, and $\ell \in \{1, \ldots, N\}$. The partial derivatives of (4) with respect to $a_0, a_1, \ldots, a_n$ are then set to zero and rearranged in order to obtain the following matrix form of the solution:

$$\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} = \begin{bmatrix} \rho_1^0 & \cdots & \rho_1^n \\ \rho_2^0 & \cdots & \rho_2^n \\ \vdots & \ddots & \vdots \\ \rho_N^0 & \cdots & \rho_N^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}. \tag{7}$$

Having estimated the parameters a_j, $j \in \{1, \ldots, n\}$ using (7), one can then estimate the distance $\hat{d}_i$ separating the stationary sensor at position $\boldsymbol{s}$ from the sensor at unknown position $\boldsymbol{x}_i$ using (6).

Remark 1. It is interesting to notice that replacing $\log_{10} d_i$ by d_i in (3) leads to a particular case of (6), where the parameters a_j, $j \in \{2, \ldots, n\}$, are null. In fact, as already mentioned, it is proven that the received signal strength decreases logarithmically with distance; in other words, the distance in logarithm scale is a linear function of the received signal strength as shown in (3).

III. PROPOSED SEMIPARAMETRIC MODEL

As we already stated, the received signal strength decreases logarithmically with distance. However, the log-distance signal propagation model of (1) alone is insufficient to characterize the RSSI/distance relationship, since it assumes that the latter is fully parametric, neglecting then all other factors in the environment, such as physical obstacles, multipath propagation, additive noises, etc. In this section, we propose a new model to characterize the relationship between the RSSIs and the distances. Our model is a semi-parametric one that combines the theoretical log-distance signal propagation model with a nonlinear fluctuation term, that represents a combination of all unknown factors affecting the RSSI measures. This term should provide the log-distance physical model with flexibility, resulting in a more accurate model.

Let $\psi(\cdot)$ denote the model that associates to each RSSI ρ_i the logarithm of the distance $\log_{10}(d_i)$. According to the definition given in the beginning of this section, $\psi(\cdot)$ can be decomposed into twos terms: a linear term and a nonlinear fluctuation term, as follows:

$$\psi(\cdot) = \psi_{lin}(\cdot) + \psi_{nlin}(\cdot) \tag{8}$$

To determine $\psi(\cdot)$, we consider the log-distance model of (1), and add a noise term φ. Then the relationship between the RSSIs and the distances is given by:

$$\rho_i = \rho_0 - 10\, n_P \, \log_{10} \frac{d_i}{d_0} + \varphi. \tag{9}$$

This model can be written as follows:

$$\log_{10} d_i = \frac{\rho_0}{10\, n_P} + \log_{10} d_0 - \frac{\rho_i}{10\, n_P} + \frac{\varphi}{10\, n_P}. \tag{10}$$

One can see that (10) is a combination of a linear model in terms of ρ_i and a nonlinear model. Thus, one can conclude the following:

$$\begin{cases} \psi_{lin}(\rho_i) = \alpha\, \rho_i + \beta, \\ \psi_{nlin}(\rho_i) = \frac{\varphi}{10\, n_P}, \end{cases} \tag{11}$$

where $\alpha = -\frac{1}{10\, n_P}$ and $\beta = \frac{\rho_0}{10\, n_P} + \log_{10} d_0$. As for the nonlinear term $\psi_{nlin}(\cdot)$, we assume that it lies in a reproducing kernel Hilbert space denoted by $\mathcal{H}_{nlin}$, and generated by a positive definite kernel function $\kappa_{nlin}(\cdot,\cdot)$. From the representer's theorem [17], [18], $\psi_{nlin}(\cdot)$ can be written as a linear combination of kernels:

$$\psi_{nlin}(\cdot) = \sum_{\ell=1}^{N} \gamma_\ell \, \kappa_{nlin}(\boldsymbol{\rho}_\ell, \cdot), \tag{12}$$

where $\kappa_{nlin} : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$, and $\gamma_\ell, \ell \in \{1, \ldots, N\}$, are parameters to be estimated. According to [19], it can be shown that the direct sum of $\mathcal{H}_{lin}$ and $\mathcal{H}_{nlin}$ of the RKHS of kernels $\kappa_{lin}(\rho_u, \rho_v) = \rho_u^\top \rho_v$ and κ_{nlin} is also a RKHS with the following kernel function:

$$\begin{aligned} \kappa(\rho_u, \rho_v) &= \rho_u^\top \rho_v + \kappa_{nlin}(\rho_u, \rho_v), \\ &= \kappa_{lin}(\rho_u, \rho_v) + \kappa_{nlin}(\rho_u, \rho_v). \end{aligned}$$

TABLE I: Some commonly used reproducing kernels, with parameters $c, \sigma > 0$, and $q \in \mathbb{N}_+$.

Type	General form
Polynomial	$\kappa_{nlin}(\rho_u, \rho_v) = \left(c + \rho_u^\top \rho_v\right)^q$
Exponential	$\kappa_{nlin}(\rho_u, \rho_v) = \exp\left(\frac{1}{\sigma} \rho_u^\top \rho_v\right)$
Gaussian	$\kappa_{nlin}(\rho_u, \rho_v) = \exp\left(-\frac{1}{2\sigma^2} \lVert\rho_u - \rho_v\rVert^2\right)$

Table I shows some of the most commonly used kernel functions κ_{nlin}.

Now the $N+2$ parameters, α, β, and $\boldsymbol{\gamma} = (\gamma_1 \ldots \gamma_N)^\top$, to be estimated can be found using the training set and the Least-Squares (LS) method. They must be chosen in a way to minimize the mean squared error on the training set, given by:

$$\frac{1}{N} \sum_{\ell=1}^{N} \varepsilon_\ell^2 + \eta \lVert \psi_{nlin} \rVert^2_{\mathcal{H}_{nlin}}. \tag{13}$$

Here, $\varepsilon_\ell = \log_{10} d_\ell - \alpha\, \rho_i - \beta - \sum_{j=1}^{N} \gamma_j \, \kappa(\rho_j, \rho_\ell)$, $\ell \in \{1, \ldots, N\}$, and the quantity η is a regularization parameter that controls the tradeoff between the training error and the complexity of the solution.

By multiplying (13) by N for convenience and writing it in matrix form, one gets the following:

$$\begin{aligned} &\boldsymbol{D}^\top \boldsymbol{D} + \alpha^2 \, \boldsymbol{\rho}^\top \boldsymbol{\rho} + N\beta^2 + \boldsymbol{\gamma}^\top \boldsymbol{K}_{nlin}^\top \boldsymbol{K}_{nlin} \boldsymbol{\gamma} \\ &- 2\alpha \boldsymbol{D}^\top \boldsymbol{\rho} - 2\beta \boldsymbol{1}^\top \boldsymbol{D} - 2\boldsymbol{D}^\top \boldsymbol{K}_{nlin} \boldsymbol{\gamma} + 2\alpha\beta \boldsymbol{1}^\top \boldsymbol{\rho} \\ &+ 2\alpha \boldsymbol{\rho}^\top \boldsymbol{K}_{nlin} \boldsymbol{\gamma} + 2\beta \boldsymbol{1}^\top \boldsymbol{\gamma} \\ &+ \eta\, N \boldsymbol{\gamma}^\top \boldsymbol{K}_{nlin} \boldsymbol{\gamma}, \end{aligned} \tag{14}$$

where $\boldsymbol{K}_{nlin}$ is the $N \times N$ matrix whose (u, v)-th entry is $\kappa_{nlin}(\rho_u, \rho_v)$, for $u, v \in \{1, ..., N\}$.

Then, the partial derivatives in matrix form of (14) with respect to α, β and $\boldsymbol{\gamma}$ are given by the following:

$$\begin{cases} \frac{1}{2} \frac{\partial}{\partial \alpha} = \boldsymbol{\rho}^\top \boldsymbol{\rho}\, \alpha + \boldsymbol{1}^\top \boldsymbol{\rho}\, \beta + \boldsymbol{\rho}^\top \boldsymbol{K}_{nlin} \boldsymbol{\gamma} - \boldsymbol{D}^\top \boldsymbol{\rho}, \\ \frac{1}{2} \frac{\partial}{\partial \beta} = \boldsymbol{1}^\top \boldsymbol{\rho}\, \alpha + N\beta + \boldsymbol{1}^\top \boldsymbol{K}_{nlin} \boldsymbol{\gamma} - \boldsymbol{1}^\top \boldsymbol{D}, \\ \frac{1}{2} \frac{\partial}{\partial \boldsymbol{\gamma}} = \boldsymbol{\rho}\, \alpha + \boldsymbol{1}\, \beta + (\boldsymbol{K}_{nlin} + \eta N \mathbf{I}) \boldsymbol{\gamma} - \boldsymbol{D}, \end{cases} \tag{15}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix.

Setting the derivatives in (15) to zero will lead to a linear system having the form $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{X}$, where

$$\boldsymbol{X} = \begin{bmatrix} \alpha \\ \beta \\ \boldsymbol{\gamma} \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} \boldsymbol{1}^\top \boldsymbol{D} \\ \boldsymbol{D} \\ \boldsymbol{D}^\top \boldsymbol{\rho} \end{bmatrix},$$

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{1}^\top \boldsymbol{\rho} & N & \boldsymbol{1}^\top \boldsymbol{K}_{nlin} \\ \boldsymbol{\rho} & \boldsymbol{1} & \boldsymbol{K}_{nlin} + \eta N \mathbf{I} \\ \boldsymbol{\rho}^\top \boldsymbol{\rho} & \boldsymbol{1}^\top \boldsymbol{\rho} & \boldsymbol{\rho}^\top \boldsymbol{K}_{nlin} \end{bmatrix} \tag{16}$$

The solution is then given by:

$$\boldsymbol{X} = (\boldsymbol{A}^\top \boldsymbol{A})^{-1} \boldsymbol{A}^\top \boldsymbol{B}. \tag{17}$$

After computing the model's parameters using (17), one can find the logarithm of any distance separating a stationary sensor and a moving sensor in the network using only the RSSI ρ_i, as follows:

$$\log_{10} d_i = \psi(\rho_i) = \alpha\rho_i + \beta + \sum_{\ell=1}^{N} \gamma_\ell \, \kappa(\rho_\ell, \rho_i). \tag{18}$$

The distance is then given by $d_i = 10^{\psi(\rho_i)}$.

Remark 2. It is interesting to see that the solution obtained in (5) is actually included in the final solution (16) of the proposed semiparametric model. Indeed, the main difference between the proposed model and the log-distance model lies in the nonlinear fluctuation term, whose inclusion induces the introduction of $\boldsymbol{\gamma}$, thus the incorporation of N new equations to the system.

Remark 3. In this paper, we use the gaussian kernel of Table I. Nevertheless, let us consider the polynomial kernel and study the effect of such choice on our model. From (18) and Table I, one can write the following:

$$\log_{10} d_i = \alpha\rho_i + \beta + \sum_{\ell=1}^{N} \gamma_\ell \, \left(c + \rho_\ell^\top \, \rho_i\right)^q, \tag{19}$$

where $c > 0$, and $q \in \mathbb{N}_+$. For $q = 1$, the terms of (19) can be rearranged as follows:

$$\log_{10} d_i = \beta + c\sum_{\ell=1}^{N} \gamma_\ell + (\alpha + \sum_{\ell=1}^{N} \gamma_\ell \rho_\ell)\rho_i.$$

This equation takes the same form as (3). Therefore, when choosing the polynomial kernel with $q = 1$, the obtained model is equivalent to the log-distance propagation model.

Now let $q = 2$; equation (19) can then be written as follows:

$$\log_{10} d_i = \beta + c^2 \sum_{\ell=1}^{N} \gamma_\ell + (\alpha + 2c\sum_{\ell=1}^{N} \gamma_\ell \rho_\ell)\rho_i + (\sum_{\ell=1}^{N} \gamma_\ell \rho_\ell^2)\rho_i^2.$$

This equation is of the following form: $\log_{10} d_i = a_0 + a_1\rho_i + a_2\rho_i^2$. For larger values of q, one will eventually be able to rearrange the terms and write the following:

$$\log_{10} d_i = a_0 + a_1 \, \rho_i + a_2 \, \rho_i^2 + \cdots + a_q \, \rho_i^q, \tag{20}$$

where a_j, $j \in \{0, \ldots, q\}$, are coefficients that are functions of α, β, c, and γ_ℓ, $\ell \in \{1 \ldots N\}$. Finally, one can conclude that choosing the polynomial kernel in our model leads to (20), which is the same model described in *Remark* 1.

IV. Experimental results

We now propose to evaluate the accuracy of the proposed model, when used for distance estimation, in the case of real data and simulated data. In the first subsection, a set of collected measurements available from [20] is used for the evaluation, and the results are then compared to ones obtained using the two state-of-the-art models of Section II. Then, in the second subsection, the model is evaluated on simulated data; the results are also compared to ones obtained using the models of Section II.

A. *Evaluation of the model on real data*

In this subsection, we use the set of collected measurements available from [20] for the evaluation of the proposed model. The measurements are carried out in a room of approximately $10m \times 10m$, where 48 uniformly distributed EyesIFX sensor nodes are deployed. Furniture and people in the room cause multi-path interferences affecting the collected RSSI values. We consider that there are 4 fixed stationary sensors at known positions, and 44 other sensors with known positions for the training and test phases. The left plot of Fig. 1 shows the topology of the testbed.

It is important to note that the average values over time of the RSSIs are used in this section. In fact, the RSSIs vary a lot with respect to time and movements, as one can see in the right plot of Fig. 1. These variations are known as short-term or multi-path fading. On the other hand, the local average of the signal varies slowly. These slow fluctuations depend mostly on environmental characteristics, and they are known as long-term fading. Therefore, it is more suitable to use the average values of the RSSIs than to use all the collected values [9].

Based on the described scenario, one can see that there are 4 signal propagation models to be determined, i.e., one model per stationary sensor. Each model has a different set of training data to be used, and different parameters that need to be found. In fact, the RSSIs of the signals exchanged between each stationary sensor and the 44 other sensors are used along with the distances separating this stationary sensor from the other sensors. This information is then used in the training phase as described in III to compute the model's parameters. We use the Gaussian kernel from Table I, whose parameters are chosen in a way to minimize the error on the training set. The value of this error for the 4 computed models is given in Table II, along with the error on the training set for the physical log-distance propagation model used in [2], [3], [4] and for the polynomial model used in [7], [8]. We also use the leave-one-out (LOO) technique in order to evaluate the performance of the proposed model in the case of data that are not part of the training set. The LOO technique involves using a single observation from the collected data as the validation data, and the remaining observations as the training data. This is repeated 44 times, such that each observation is used once as the validation data. This technique is interesting because it allows us to compare our proposed model to the log-distance model and the polynomial model, even though the set of collected data is not really large. Finally, the mean estimation error averaged over the 44 simulations is stored in Table II for the 4 computed models, in the case of the proposed model, as well as the two already-described ones. One can see from Table II that the proposed model yields better results than the state-of-the-art models, when comparing the mean estimation error.

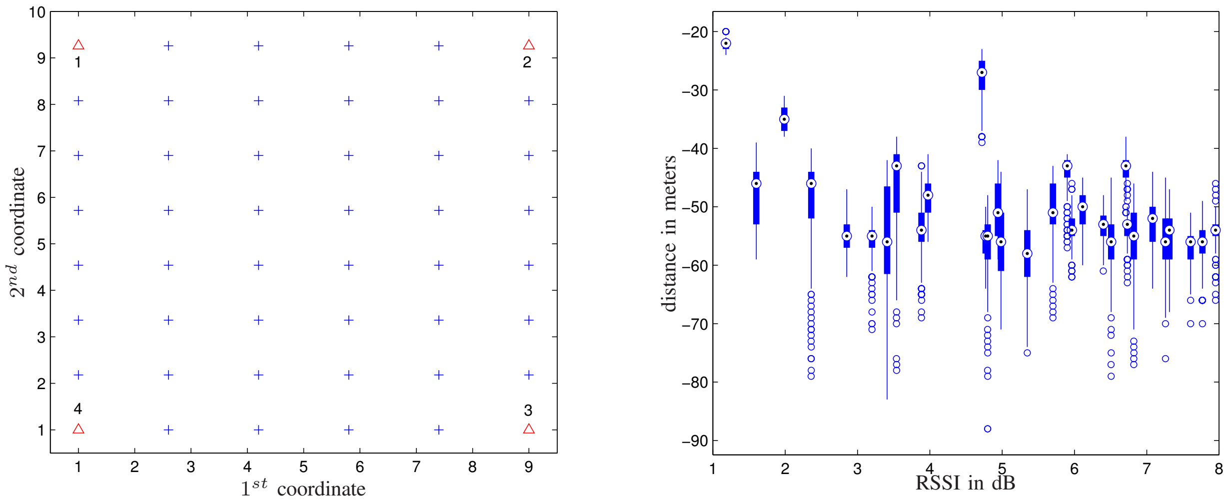

Fig. 1: Left plot: Topology of the testbed, where Δ represents the stationary sensors and $+$ represents the training positions. Right plot: RSSIs measured at the training positions as a function of the distances separating these positions from the stationary sensor 1.

TABLE II: Comparison between models (error in meters) for real data

		Physical model	Polynomial model ($n = 2$)	Polynomial model ($n = 3$)	Polynomial model ($n = 4$)	**Proposed model**
Model 1	Training error	1.14	1.13	1.09	1.10	0.44
	LOO error	1.19	1.24	1.21	1.76	1.07
Model 2	Training error	1.55	1.41	1.40	1.40	1.42
	LOO error	1.62	1.48	1.51	1.54	1.51
Model 3	Training error	1.47	1.44	1.44	1.41	1.32
	LOO error	1.52	1.55	1.88	2.53	1.46
Model 4	Training error	1.66	1.65	1.65	1.62	1.47
	LOO error	1.72	1.78	1.78	1.83	1.71
Mean error	Training error	1.46	1.41	1.39	1.38	**1.16**
	LOO error	1.51	1.51	1.60	1.91	**1.43**

B. Evaluation of the model on simulated data

In this subsection, we evaluate the proposed model using simulated data. To this end, we consider the average walls model described in [11] to generate the RSSI measures. This model is a modified version of the log-distance model that explicitly takes into account the attenuations due to walls. The received signal strength indicator is then given by the following:

$$\rho_i = \rho_0 - 10\, n_P \, \log_{10} \frac{d_i}{d_0} - N_{w_i} L_{w_i} + \epsilon_i, \qquad (21)$$

where ρ_i (in dB) is the power received from the stationary sensor at position $\boldsymbol{s}$ by the sensor at position $\boldsymbol{x}_i$, ρ_0 is the power at the reference distance d_0 set to $1dB$, while the path-loss exponent n_P is set to 4. As for the quantities L_{w_i} and N_{w_i}, they denote respectively the loss due to walls and the number of penetrated walls. The quantity L_{w_i} is taken equal to $6.9dB$, since we consider the case of heavy thick walls [11]. Finally, the quantity ϵ_i is a zero mean additive white noise, whose standard deviation is taken equal to 1% of the standard deviation of the RSSIs.

Now consider the $25m \times 5m$ area given in Fig. 2, where 2 fixed stationary sensors and 45 known positions for the training phase are considered. This figure shows that there are 5 rooms in the area; therefore, the signal penetrates a maximum of 4 walls during its propagation, i.e., $N_{w_i} \in \{0 \dots 4\}$. As for the test phase, 100 positions are randomly

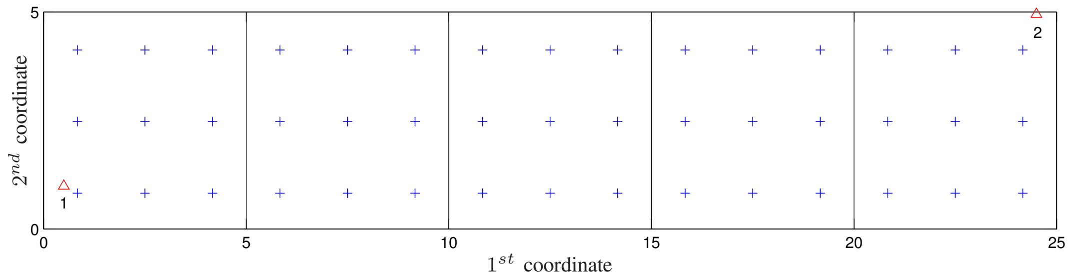

Fig. 2: Topology of the simulated area, where Δ represents the stationary sensors and $+$ represents the training positions.

TABLE III: Comparison between models (error in meters) for simulated data

		Physical model	Polynomial model ($n = 2$)	Polynomial model ($n = 3$)	Polynomial model ($n = 4$)	**Proposed model**
Model 1	Training error	1.36	0.55	0.51	0.52	0.21
	Test error	1.25	0.60	0.57	0.57	0.23
Model 2	Training error	1.13	0.57	0.55	0.55	0.17
	Test error	1.17	0.58	0.57	0.58	0.22
Mean error	Training error	1.25	0.56	0.53	0.54	**0.19**
	Test error	1.21	0.59	0.57	0.58	**0.23**

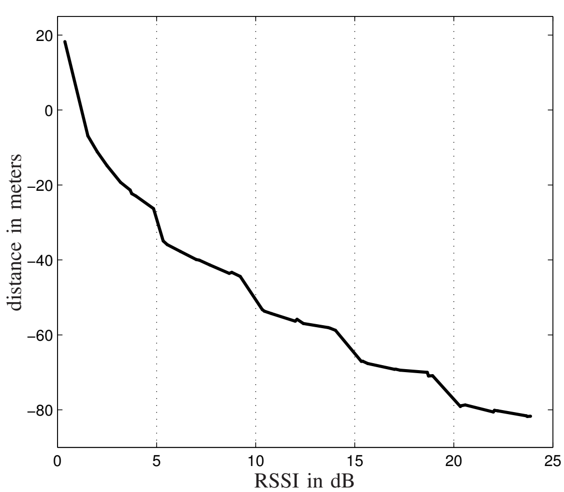

Fig. 3: RSSIs measured at the training positions as a function of the distances separating these positions from the stationary sensor 1 of Fig. 2 - Simulated data.

generated in the studied area, and their RSSIs are obtained using (21). Fig. 3 shows the RSSIs measured at the 45 training positions along with the distances separating these positions from the stationary sensor 1 of Fig. 2. One can see that the signal decreases with the traveled distances, and that the loss gets higher every time the signal penetrates a wall.

We now have to determine the two signal propagation models' parameters. As we previously explained, the RSSIs of the signals and the distances are used to estimate these parameters. The Gaussian kernel from Table I is used here as well. Table III shows the errors on the distances that are estimated using the proposed model, the physical log-distance propagation model, and the polynomial model. Compared to the estimation error obtained using the state-of-the-art models, the estimation error is reduced by half or more when using the proposed model, proving that the latter outperforms the other models in terms of accuracy.

V. Conclusion

In this paper, we proposed a semiparametric regression model that combines the well-known log-distance propagation model with a nonlinear fluctuation term, estimated within the framework of kernel-based machines. Evaluation on real and simulated data showed that the proposed model outperforms state-of-the-art models in terms of accuracy. Future works will handle improvements of the quality of the RSSIs; for instance, a filtering process can be considered to reduce the noise on the RSSIs in the case of real data before finding the signal propagation model.

REFERENCES

[1] E.-E.-L. Lau and W.-Y. Chung, "Enhanced RSSI-based real-time user location tracking system for indoor and outdoor environments," in *International Conference on Convergence Information Technology*, 2007, pp. 1213–1218.

[2] A. Medeisis and A. Kajackas, "On the use of the universal Okumura-Hata propagation prediction model in rural areas," *Vehicular Conference Proceedings*, vol. 3, 2000.

[3] N. Patwari, J. Ash, S. Kyperountas, A. Hero, R. Moses, and N. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 54–69, July 2005.

[4] A. Zanella and A. Bardella, "Rss-based ranging by multichannel rss averaging," *IEEE Wireless Communications Letters*, vol. 3, no. 1, pp. 10–13, February 2014.

[5] S. Seidel and T. Rappaport, "914 mhz path loss prediction models for indoor wireless communications in multifloored buildings," *IEEE Transactions on Antennas and Propagation*, vol. 40, no. 2, pp. 207–217, Feb 1992.

[6] A. Sandeep, Y. Shreyas, S. Seth, R. Agarwal, and G. Sadashivappa, "Wireless network visualization and indoor empirical propagation model for a campus wi-fi network," *World Academy of Science, Engineering and Technology*, vol. 2, no. 6, pp. 706 – 710, 2008.

[7] Y. Wang, X. Jia, H. K. Lee, and G. Y. Li, "An indoors wireless positioning system based on wireless local area network infrastructure," *Technology Including Mobile Positioning*, 2003.

[8] J. Yang and Y. Chen, "Indoor localization using improved rss-based lateration methods," in *IEEE Global Telecommunications Conference GLOBECOM.*, Nov 2009, pp. 1–6.

[9] A. Neskovic, N. Neskovic, and G. Paunovic, "Modern approaches in modeling of mobile radio systems propagation environment," *IEEE Communications Surveys Tutorials*, vol. 3, no. 3, pp. 2–12, Third 2000.

[10] T. Sarkar, Z. Ji, K. Kim, A. Medouri, and M. Salazar-Palma, "A survey of various propagation models for mobile communication," *IEEE Antennas and Propagation Magazine*, vol. 45, no. 3, pp. 51–82, June 2003.

[11] C. Andrade and R. Hoefel, "IEEE 802.11 WLANs: A comparison on indoor coverage models," in *Electrical and Computer Engineering (CCECE), 2010 23rd Canadian Conference on*, May 2010, pp. 1–6.

[12] S. Mahfouz, F. Mourad-Chehade, P. Honeine, H. Snoussi, and J. Farah, "Kernel-based localization using fingerprinting in wireless sensor networks," in *IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2013, pp. 744–748.

[13] ——, "Decentralized localization using fingerprinting and kernel methods in wireless sensor networks." in *Proc. 21th European Conference on Signal Processing (EUSIPCO)*, 2013.

[14] S. Mahfouz, F. Mourad-Chehade, P. Honeine, J. Farah, and H. Snoussi, "Target tracking using machine learning and kalman filter in wireless sensor networks," *Sensors Journal, IEEE*, 2014.

[15] D. Liu, X. Lin, and D. Ghosh, "Semiparametric regression of multidimensional genetic pathway data: Least-squares kernel machines and linear mixed models," vol. 63, pp. 1079–1088, 2007.

[16] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 480–492, Jan 2013.

[17] G. Kimeldorf and G. Wahba, "Some results on tchebycheffian spline functions," *Journal of Mathematical Analysis and Applications*, vol. 33, pp. 82 – 95, 1971.

[18] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *Proc. of the 14th Annual Conference on Computational Learning Theory and 5th European Conference on Computational Learning Theory*. London, UK: Springer-Verlag, 2001, pp. 416–426.

[19] D. Haussler, "Convolution kernels on discrete structures," University of California at Santa Cruz, Technical Report UCS-CRL-99-10, 1999.

[20] G. Zanca, F. Zorzi, and A. Zanella, "RSSI measurements in indoor wireless sensor networks." [Online]. Available: http://telecom.dei.unipd.it/pages/read/59/