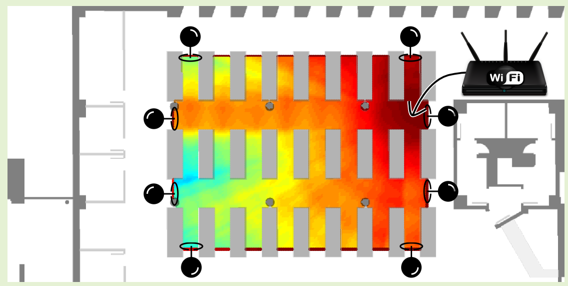
Environment-Aware Regression for Indoor Localization based on WiFi Fingerprinting

Germán Mendoza-Silva, Ana Cristina Costa, Joaquín Torres-Sospedra, Marco Painho and Joaquín Huerta

***Abstract*—Data enrichment through interpolation or regression is a common approach to deal with sample collection for Indoor Localization with WiFi fingerprinting. This paper provides guidelines on where to collect WiFi samples , and proposes a new model for received signal strength regression. The new model creates vectors that describe the presence of obstacles between an access point and the collected samples. The vectors, the distance between the access point and the positions of the samples, and the collected, are used to train a Support Vector Regression. The experiments included some relevant analyses and showed that the proposed model improves received signal strength regression in terms of regression residuals and positioning accuracy.**



***Index Terms*—Indoor Positioning, WiFi Fingerprinting, WiFi Samples Collection, RSS Regression**

I. Introduction

The demand for Indoor Positioning Systems (IPS) has already driven academic and commercial research, it is expected that it will dramatically rise in the years to come [1]. Despite the large diversity on related positioning technologies for indoor scenarios, WiFi is one of the most often used. Smartphones and applications relying on Location Based Services (LBS) made WiFi a cost-less approach at the expense of positioning errors around a few meters [2].

Fingerprinting is commonly used with WiFi to provide position indoors. A WiFi fingerprint is a vector with the Received Signal Strength (RSS) of each WiFi access point (AP) detected in a given position and time. It requires a calibration stage, where samples are collected at well-known positions to create a reference dataset (radio map). In the operational stage, given a new fingerprint measured at an unknown position, the fingerprint method usually provides the centroid of the most similar reference fingerprints as position estimate [3].

Samples collection is known as one of the main challenges of WiFi fingerprinting [4], given that the collection effort can be significant for large areas. The literature suggests to reduce the required effort either by crowdsourcing the collection to volunteers [5], estimating the RSS values applying a propagation model, or applying an interpolation technique to densify an initial reduced radio map [4], [6], [7]. Despite being very valuable, the reliability of position tags and the improper distribution sample position are usual concerns with crowdsourced signal data [8].

This paper addresses the radio map enrichment by applying regression techniques on a proper signal characterization of the environment. Also, through experiments performed on two publicly available databases, we address the problem of choosing the most convenient positions for collecting WiFi fingerprints for radio map creation. Furthermore, we evaluate a new model that applies environment knowledge to Support Vector Regression (SVR), which improves the regression estimates corresponding to extrapolation points in comparison to other extrapolation work shown in WiFi positioning literature.

The main contributions of this paper can be summarized as follows: i) a novel regression model aware of the environment features; ii) a comprehensive analysis of reference position selection to build effective radio maps; and iii) validation in a real-world scenario independent to the research objectives.

II. Background and Related Works

A WiFi Access Point (AP) is a networking device that broadcasts one or more wireless networks. A set of RSS values from available APs measured at a specific location throughout a short time interval is called a fingerprint, which can be used for positioning as described in Section I. The quality of the radio map depends on the location of the reference points, the reference point density, the number of samples of each reference points, among many other parameters [9], [10].

Manuscript received xx December 2019; revised xx April 2020; accepted xx xxxx 2020. Date of publication xx xxxx 2020; (Corresponding author: J. Torres-Sospedra.). G. M. Mendoza-Silva gratefully acknowledges funding from grant PREDOC/2016/55 by Universitat Jaume I. J. Torres-Sospedra gratefully acknowledges funding from Ministerio de Ciéncia, Innovación y Universidades (PTQ2018-009981).

G. M. Mendoza-Silva and J. Huerta are with the Institute of New Imaging Technologies, Universitat Jaume I, 12071 Castellón, Spain. (e-mail: {gmendoza,huerta}@uji.es)

J. Torres-Sospedra (corresponding author), is with UBIK Geospatial Solutions, 12006 Castellón, Spain. (e-mail: torres@ubikgs.com)

A. C. Costa and M. Painho are with the NOVA IMS, Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal. (e-mail: {cristina,painho}@@novaims.unl.pt)

Corresponding author: J. Torres-Sospedra (torres@ubikgs.com)

Digital Object Identifier: 10.1109/JSEN.202x.xxxxxxx

However, collecting samples for a radio map requires a notable amount of time [11]. To tackle this problem, two alternatives are usually considered: crowdsourcing and sparse collection. Crowdsourcing has been praised for radio map collection and update [12] at the expense of suffering from low quality of position tags or uneven distributions of the collected samples, whereas sparse collection reduces the collection efforts at the expense of poorer characterizations of the environments. The later approach (sparse collection with regression, interpolation and/or extrapolation models) has been applied to synthetically enrich the radio map for more than 15 years [13], [14], and methods fall in one of the next groups:

- *Sparse recovery* includes, for example, compressed sensing using Singular Value Decomposition (SVD) [15], and radio map interpolation using sparse recovery [6].
- *Interpolation methods* includes traditional interpolation methods [16]–[19]; methods capable of delivering both interpolation and extrapolation like Nearest Neighbor and Inverse Distance Weighting (IWD) [20]; and other interpolation heuristics [21].
- *Extrapolation methods* applied variants based on log-distance path loss model [21]–[23]; on the ray tracing model [24], [25] or the radiosity model [26]–[28].
- *Regression methods* largely includes the application of Gaussian Process Regression (GPR) [29]–[33], although others have also applied Kriging [14], [34]–[36], Geography Weighted Regression (GWR) [37] and Support Vector Regression (SVR) [38].

It is common that radio map enrichment works provide the proportions between points used for fitting and those used for estimations. Talvitie *et al.* [20] concluded that the positions where samples are selected were more important than how many of them were selected. Khalajmehrabadi *et al.* [6] suggested a random selection of reference points and discourage a uniform placement of those points. Ezpeleta *et al.* [16] supported the division in zones arguing that a zone with higher quality of RF signals than other zones required less training points. The importance of the distribution of samples for radio map construction is almost intuitive and acknowledged [39]. However, some works perform random selection of sample positions for radio map construction [6], [23], [32]. Kanaris *et al.* [40], determined the sample size given a small preliminary set of measurements, suggesting to randomly choose positions from a grid in the number determined by the sample size calculation.

Some radio map enrichment solutions have considered the environment's influence on the signals intensities. The interpolation in Bong *et al.* [41] preserved signals discontinuity over the wall. Ali *et al.* [23] used a path loss with wall attenuation factor that introduced an image to count the number of interfering walls. Moghtadaiee *et al.* [21] fitted a log-distance model independently for each architectural zone and created an interpolation that considered only sample at similar distances to the target AP. Some authors [14], [34]–[36] used Kriging, but only considered the Euclidean distance for describing the spatial dependency, which does not hold true for indoor environments. [39] fitted a log distance path loss model for each target position, giving to the samples used for fitting distinct weights (using a kernel density estimation) based on their distances to the target position. Du *et al.* [37] applied GWR, which computed several local models instead a single global one. They used the distance between the emitters and the sample points as predictor variables.

The distribution of samples necessarily should take the layout of the environment into account, not only regarding where it is possible to collect samples, but where is convenient to collect them. The indoor environments strongly influence the WiFi and BLE signals, and the decision on the collection distribution should be aware of it. The radio map enrichment method should ideally be also aware of the target environment, i.e., of the obstacles and the positions of the emitters.

III. MATERIALS AND METHODS

A. Selected datasets

This work is built on top of two public WiFi fingerprinting datasets: the Library dataset [42] and the Mannheim dataset [43]. Partial versions of both datasets will be used to analyse the influence of position distribution and the influence of AP strength on position accuracy. Moreover, they will be used to analyse the influence of AP strength on RSS regressions. For the evaluation of our proposed environment-aware regression model only the Library dataset will be used.

The Library dataset was collected in two floors of the Library building of University Jaume I (UJI) and the systematic data collection was repeated multiple times in a time span of 25 months. There are six WiFi fingerprints per each reference point and each of the two directions at which the collection subject was facing. Also, as the data contained information about a 620 AP, a selection of the 52 most relevant APs was performed (as done in Torres-Sospedra *et al.* [44]) to ease the analyses and and reduce the noise created by a large number of intermittent APs. The collection area is a relatively small environment that covered about $15 \times 10\,\mathrm{m}$. The average distance between reference points is about $2\,\mathrm{m}$.

The Mannheim dataset was collected in the Mannheim University. The collection area comprises a medium-scale environment, covering about $50 \times 36\,\mathrm{m}$ of corridors of a university department. The fingerprints are on a $1.5\,\mathrm{m}$ grid [43], [45] and the positions of 10 APs are known. The dataset contained 110 fingerprints per reference point. Out of 110 samples, we randomly selected 10 to ease the analyzes and have a number of samples that is closer to that of the Library dataset. Both the original Mannheim and the Library datasets provided their position tags using a local coordinate system that allows distance computation using the Euclidean distance.

Figure 1 shows the operational area of the two evaluation environments. The structural barriers were manually created from floor plans. Thick walls were drawn in black color and thin walls were drawn with a light shade of gray in the image, whose intensity values are used by eq.(2). Figure 1 also presents the distribution of training and test reference points, as well as the position of some APs. The higher the density of APs and reference points in the operational area, the lower expected positioning error. In both cases, some APs lay out the floormap or have an unknown location.

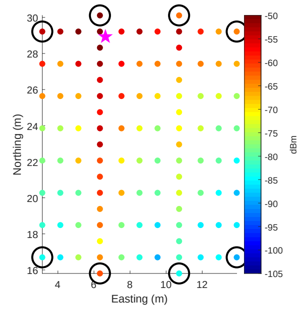
Fig. 1: Library 5^{th} floor (left) and Mannheim (right) floormaps. Blue and magenta dots represent training and test reference points, respectively. APs positions are drawn with orange circles. Other APs may lay out of the areas.

B. Environment-Aware Regression on WiFi radio maps

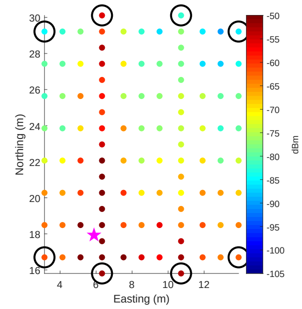
This work presents a regression model that integrates the AP reference position and a floor plan of the area. The reference position is used as a raw indication of where the AP is. The position of APs inside or very close to the collection area can be determined with, for instance, the weighted centroid or the method proposed in [46]. The approximate position of an AP can be also manually obtained by measuring the signal intensity with a smartphone application walking in the area. However, the accuracy for AP location is low for those APs that are away from the operational area and an indicator of the relative direction is obtained instead. Those far APs are typically detected with a maximum intensity weaker than −60 dBm. Determining whether an AP is within the collection area could be done, for instance, using the Situation Goodness test presented in [46] if a relatively dense sample collection is available.

Figure 2 introduces an example in the Library environment (5th floor). It shows the mean RSS values per reference point for 3 APs, which will later be used to evaluate the proposed regression model. The APs with IDs 15 and 49 are inside the collection area. Their positions shown in the figure are about half a meter and more than a meter away from the actual device positions, respectively. The position of the device that emitted the AP with ID 8 was unknown. The position shown in the figure is anyway a useful estimation of the actual AP direction.

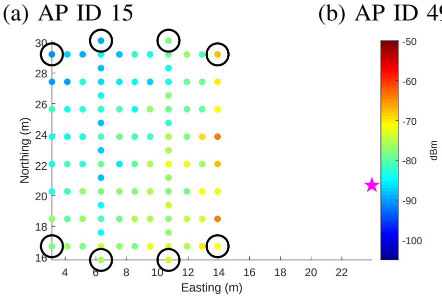
In the proposed model, the predictor variables include the target point's position components, the AP's reference position and information from the environment floor map. Moreover, we applied a data transformation before and after the application of the regression method, so that the values of the response variable are determined as $\log_{10}(-RSS)$ (as a distance indicator) and the RSS estimate is computed as $-(10^{est})$ if est is an estimate provided by the regression model. The positions of points used for training and testing the model are expressed in the local coordinate system. Thus, their coordinates need to be transformed into image coordinates (cell positions or *pixels*) before applying the proposed model. The following definitions assume positions in image coordinates (i.e. *pixels* not meters).



(a) AP ID 15



(b) AP ID 49



(c) AP ID 8

Fig. 2: Mean RSS values per reference point and device reference positions of three APs (Library, 5th floor). The device position is indicated with a star. Circles highlight the reference points whose values where used to train regression models.

Let $rp = (rp_x; rp_y)$ be the position of a reference point used for training the model. Let $ap = (ap_x; ap_y)$ be the position of the AP targeted for regression. Let $B_{rp} = \{(x_1, y_1), \ldots, (x_k, y_k)\}$ be the line that connects rp and ap. The cell positions that constitute the line are determined using the Bresenham's line algorithm [47]. The values of predictor variables for rp are:

$$P_{rp} = \{rp_x, rp_y, \frac{d_{rp} + 1}{2}, F_{rp}\}, \tag{1}$$

where $F_{rp} = \{f_1, \ldots, f_k, \ldots, f_n\}$ and d_{rp} is the Euclidean distance between rp and ap. The value f_i is computed as:

$$f_i = \begin{cases} \log_2(2 + 255 - Im(x_i, y_i)) & \text{for } 1 \leq i \leq k \\ 0 & \text{for k} < i \leq n \end{cases} \tag{2}$$

where x_i and y_i are the position components of the i^{th} point in B_{rp}, Im is the image representation of the environment, and $Im(x_i, y_i)$ is the cell value in the image Im whose position is (x_i, y_i). The value of n is the maximum number of points that may have a line connecting the positions of the AP and a point in the environment representation. If ap lies beyond the environment represented by Im, the image is enlarge applying a padding of zeros. In other words, $Im(x, y) = 0$ for all (x, y) that lies beyond the environment representation.

Algorithm 1 resumes the process of training the proposed regression model. Its inputs are the environment image Im, the positions (expressed in a local coordinate system) of collection points $RPL = \{rpl_j\}$ and their respective RSS values $SI = \{si_j\}$ measured for an AP. Once the model M is ready, it serves for predicting the RSS values $SO = \{so_j\}$ for a set of positions $TPL = \{tpl_j\}$ using the Algorithm 2.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3073878, IEEE Sensors Journal

Algorithm 1: Regression model training for an AP

Input: Im, RPL, SI, apl **Output:** The trained regression model M

Compute $ap = (ap_x, ap_y)$, the position of apl_j in Im
for *each* rpl_j *in* RPL **do**
  Get $rp = (rp_x, rp_y)$, the position of rpl_j in Im
  Get B_{rp}, as stated previously
  Get P_{rp}, as stated in Equation 1
  Set $p_j = P_{rp}$
  Get $resp_j = \log_{10}(-si_j)$
end
Build M by training SVR using $\{p_j\}$ as predictors data and $\{resp_j\}$ as responses data

Algorithm 2: Signal prediction for an AP

Input: Im, TPL, apl, M
Output: The predicted intensities SO

Compute $ap = (ap_x, ap_y)$, the position of apl_j in Im
for *each* tpl_j *in* TPL **do**
  Get $rp = (rp_x, rp_y)$, the position of tpl_j in Im
  Get B_{rp}, as stated previously
  Get P_{rp}, as stated in Equation 1
  Set $p_j = P_{rp}$
  Get est_j using M with $\{p_j\}$ as predictors values
  Set $so_j = -(10^{est_j})$
end
Set $SO = \{so_j\}$

The set F_{rp} in Equation 1 is a representation of the obstacles between rp and ap using the information of the image's cells that lie in that path. The cell values in the image Im represent either free space or an obstacle (black or white). Thus, the model is trained to learn the influence of an obstacle cell value at a given distance from an AP in the signal propagation. This work did not differentiate among distinct types of obstacle materials for simplicity, despite Equation 2 allows the range $[1, \ldots, 255]$ for obstacle representation. Setting appropriate opaqueness for each material requires additional consideration and measurements. Equation 1 includes half of the distance between rp and ap. Using the actual value of the distance significantly decreased the obstacles influences in the model. The number of variables presented in Equation 1 depends on the environment and the AP position. Finally, according to our experience, we selected the Support Vector Regression (SVR) with a linear kernel function as regressor.

IV. Experiments and Results

A. Influence of RPs Distribution on IPS Accuracy

The goal of the radio map in WiFi fingerprinting is to characterize the signal propagation in the target environment. As the main fingerprint methods (including k-NN) can only provide position estimates within the convex hull of the reference sample locations, we hipotetise that the number and distribution of the collected samples are strongly related quality of the radio map and, hence, the accuracy of the IPS.

For that purpose, we have evaluated the performance of the radio map in two environments and four different cases: with 100%, 75%, 50% and 25% of RPs. Except for 100%, we repeated the evaluation 400 times with different initialization to cover multiple random scenarios. In all cases, we report the results provided by the optimal k-value (from the set $[1, \ldots, 15]$). The results are reported as a scatter in Figure 3 for Library 5$^{\text{th}}$ floor (left) and Mannheim (right).

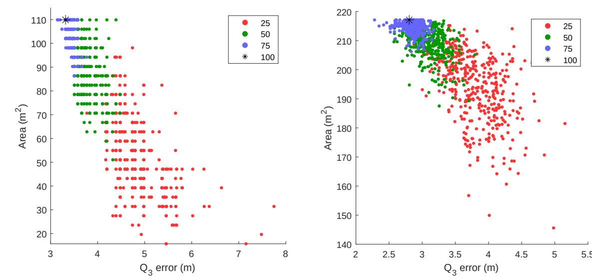


Fig. 3: Relation between training covered area and positioning accuracy for Library 5th Floor (left) and Mannheimm (right).

Every point in the figure represent the area of the reduced radio map's convex hull and the best accuracy reported by the k-NN method with that data set. The accuracy corresponds to the Q_3 value, i.e., the 75^{th} percentile as done in IPIN Competition. The color indicates the size of the radio map case (100%, 75%, 50% and 25%). A clear trend can be observed in the two environments, the large the area covered, the best positioning accuracy. In contrast, the worst positioning results came when the convex hull of the reference radio map was small. This is because the kNN method can provide position estimates only within the convex hull of the reference points. Good accuracy can be reached with a reduced radio map if the reference points cover the full operational area.

The figure also shows that the distribution of reference points is relevant. Even for a high covered area, the positioning accuracy can vary up to more than 2 m in the three cases. The largest differences in positioning are observed for cases with low RPs density (i.e. 25%). To evaluate the relation between covered area and accuracy, we calculated the Pearson correlation between the area and the Positioning error in the third quartile in the 1201 points. The correlation factor (ρ) for Mannheim is -0.77, whereas it is -0.89 for Library. In both cases, the significance (p-value) is much lower than 0.05 showing that the inverse correlation is statistically significant.

Our hypothesis is that placing reference points near the inner boundary of the collection area would maximize the covered area and assure that test positions are located inside the convex of the training positions. Finding those positions is a trivial task and can be provided by, for instance, alpha-shape [48]. Thus empirical data collection can be optimized to relevant places according to the imposed restrictions. The restrictions will somehow will be an indicator of the density and distribution of the empirical reference points, which will be located only at feasible locations (e.g. there are no samples inside a wall). If the radio map needs to be enrichted, regression can be used to synthetically generate new reference samples in those positions that lack of empirically collected data.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3073878, IEEE Sensors Journal

One strategy for creating the set of reference points is to first add reference points lying close to environment boundary and later add a number of points mp that maximize the mean minimum distance among the points in the set. In kNN, the estimated position is commonly computed as the centroid of the positions of the most similar samples in the training dataset. Thus, maximizing the minimum distance among the reference points reduces the areas without position estimates produced by kNN. Such an even distribution of point also benefits regressions as it provides intermediate positions that help explain non-linear behaviors. The value of mp may be dictated by the affordable collection effort. For low values of mp, like those below 20, a brute force approach may be applied to determine the mp positions of the reference points. For large mp values, a Monte Carlo approach [49] can be used. This work used an optimization approach based on agents moving under repulsion forces [50].

To explore the convenience of using the previous training points distribution, the Pearson correlation test was applied between the mean minimum distance and the positioning error for several distributions of training points. The tests were performed 400 independent times (with random sets of reference points that included the shape boundary) separately for each of the two environments. The position estimations were obtained with kNN, using the best k for the training set.

Table I presents the correlation results. The negative correlation between the mean minimum distance and the positioning accuracy is not statistically significant. For the Library environment, the negative weak to moderate correlation appears only for large sets, and it is statistically significant for them. The correlation is consistently negative for all set sizes in the Mannheim environment. However, its statistical significance does not show a clear pattern. The results from Table I suggest that the distribution of the inner reference points proposed above is beneficial for environments that are large or have relatively dense collections. Despite it is desirable to avoid the existence of non-positionable zones, alternative distributions may be preferable for other environments.

TABLE I: Correlation (ρ) and statistical significance (p-value) between the mean minimum distance among training points and the Third quartile of the positioning error (Q_3) for different sizes of the radio map (from 25% to 90% of RPs).

	Library				Mannheim			
%RPs	mean k	ρ	p-value	Q3 (m)	mean k	ρ	p-value	Q3 (m)
25	2	0.004	0.932	4.740	3	-0.144	0.004	3.898
30	2	0.075	0.135	4.480	3	-0.069	0.167	3.562
35	2	0.122	0.015	4.390	3	-0.120	0.016	3.403
40	2	0.091	0.068	4.171	3	-0.086	0.086	3.319
45	2	0.024	0.628	3.971	3	-0.146	0.003	3.193
50	2	0.060	0.233	3.661	3	-0.131	0.009	3.146
55	2	0.036	0.470	3.576	3	-0.130	0.010	3.071
60	3	-0.036	0.471	3.576	3	-0.077	0.122	2.990
65	3	-0.124	0.013	3.505	3	-0.085	0.088	2.966
70	3	-0.200	0.000	3.466	3	-0.075	0.134	2.926
75	4	-0.313	0.000	3.428	4	-0.152	0.002	2.900
80	4	-0.395	0.000	3.390	4	-0.138	0.006	2.874
85	4	-0.302	0.000	3.322	4	-0.056	0.267	2.864
90	4	-0.279	0.000	3.318	4	-0.131	0.009	2.833

B. Influence of AP Strength on Positioning Accuracy

It is known that the signal strength from an AP logarithmicly decreases as the distance to the AP increases. Thus, it is expected that the closer to the emitter the larger the expected variations in the signals. A radio map should grasp as much as the signal variations in the environment as possible. Having reference points close to the emitter increases the likelihood of incorporating much of those variations.

This subsection explores the correlation between AP proximity to the collection area and the positioning accuracy of a kNN method. Determining the distance to an AP requires knowing the actual position of the AP. Given that the knowledge of AP positions is commonly not assumed for fingerprinting, we inferred proximity from the RSS values. The RSS values for an AP measured in an area should be strong if the AP is close to that area or inside it.

Let us assume a radio map RM (training set) and a test set. Let $max_a = max(\{r_{p,i,a}\})$ be the strongest RSS value for the a^{th} detected AP in RM, with $1 \leq a \leq m$ and m being the number of APs. Let qap (median inferred proximity) be the Q_2 value of $\{max_a\}$. Let qpe (positioning accuracy) be the Q_3 value of positioning errors obtained by a kNN method using the above training and test sets.

Here, we also created 400 random subsets containing the 25% of an original training set (either for the Library or Mannheim). For each subset RM_s, the qap_s and qpe_s were computed. For qpe_s, the kNN method used RM_s as training set and the original test set. Then, the Pearson correlation test was applied on the sets $\{qap_s\}$ and $\{qpe_s\}$, with $1 \leq s \leq 400$. The test results are shown in Table II. For the two environments, the correlation results were statistically significant. The low to moderate negative correlation indicates that high accuracy is associated with low proximity values (weak RSS). Thus, the results suggest the convenience of distributing some reference points in zones of the collection area where nearby APs are may result in large signal variations.

TABLE II: Correlation test results between qap (median inferred proximity) and qpe (positioning accuracy).

Environment	ρ	p-value
Library	-0.37	≈ 0
Mannheim	-0.28	≈ 0

C. Influence of AP Strength on RSS Regressions

The following experiments addressed the notion of the convenience of having more reference points close to nearby APs in relation to the regression or interpolation results. The goodness of a regression or an interpolation applied to radio map densification is normally assessed by the difference between the estimated RSS and their actual values. The interpolation methods used in the experiments were Natural Neighbours [51], (Bi)Cubic Interpolation [52], [53] and Inverse Distance Weighting [54]. The regression methods used in the experiments were Support Vector Machines (SVM) [55], Gaussian Process [56], Generalized Linear Models [57],

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2021.3073878, IEEE Sensors Journal

Decision Trees (DT) [58], and Ensembles of Decision Trees [59]. The interpolation and regression methods, hereinafter only called regression methods, were applied using training points to fit the model and tests point to compute RSS estimates. The mean RSS value for an AP and a reference point was used to train the regression model for an AP and to later compute the regression residuals. The residuals are the AP-wise absolute difference between RSS estimates provided by the regression and the actual RSS used for training.

Table III shows the correlation values between signal strength and regression residuals for each environment. Let $S_j = \{s_1, \ldots, s_n\}$ and $R_j = \{r_{j,1}, \ldots, r_{j,n}\}$ be two sets, where n is the number of APs detected in that environment. The value s_i was computed as the mean RSS value of the i^{th} AP in the environment, considering all reference points. The value $r_{j,i}$ was computed as the mean of the residual values obtained for the i^{th} AP applying the j^{th} regression method in the environment. The values for the signal strength and regression residuals used for the correlation test in an environment are the sets $\{S_1, \ldots, S_m\}$ and $\{R_1, \ldots, R_m\}$, where m is the number of regression methods.

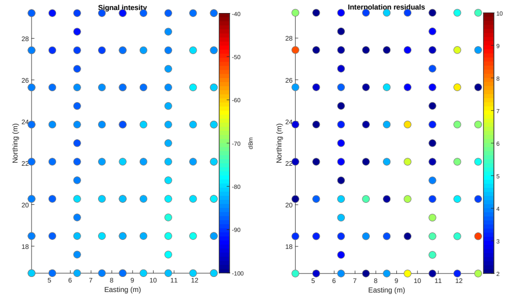
TABLE III: Correlation between mean values of signal strength in the environment and mean values of regression residuals.

Environment	ρ	p**-value**
Library	0.88	≈ 0
Mannheim	0.24	≈ 0

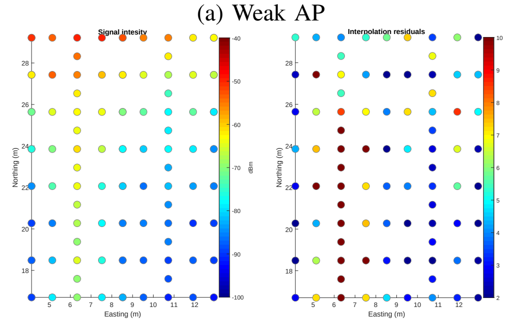
The correlation is statistically significant for the two environments. The correlation magnitude is weak for the Mannheim environment but notable for the Library environment. The higher the median value of the signal strength in the environment, the larger the residuals of the regressions. The correlation difference between the two environments is a likely result of the dimensions of the environments. The Mannheim environment is large, and thus the detected signal intensities for an AP can be very strong in some areas and very weak at some other areas. Very strong and very weak signal intensities are not detected for the same AP in the Library environment.

Figure 4 shows the relation between the strength with which an AP is seen in an environment and the regression goodness. The investigation was performed for two APs in the Library environment (one with weak and one with strong RSS values). The charts from Figure 4 present for each AP includes the median value for the RSS values of the AP at each reference point and the median value of the regression residuals at each reference point. In particular, figure 4a shows regression residuals of moderate values for the weak AP, while Figure 4b shows regression residuals for the strong AP that are not only notably larger than those for the weak AP but also mainly situated in a specific zone of the environment.

The charts suggest that for weak, far away APs, the regression requires only a few samples to train a model, as the APs signals are only weakly affected by the environment. However, the strength values of signals from APs near the target environment heavily depend on the Line of Sight (LOS) and Non Line of Sight (NLOS) situations.



(a) Weak AP



(b) Strong AP

Fig. 4: Mean value of residuals distribution compared to mean value of RSS for the Library environment.

Table IV presents the spatial auto-correlation test as obtained by Moran's I [60] for the two antennas addressed in Figure 4. The table suggests that for APs strongly seen across the environment the distribution of regression residuals is not random and tends to organize in clusters; while for APs weakly seen in the environment the distribution of residuals is likely random. As stated in the literature [61], the environment influence is less significant for weak than for strong signals. Furthermore, the signal in free space follows a logarithmic decay, i.e., the farther from the AP the slower the decay rate. The tested regression models fail to account for a spatial process induced by the environment for strong signals. Thus, samples are required in zones of LOS and NLOS with respect to nearby APs, given that the RSS values in those two situations can be significantly different.

TABLE IV: Spatial auto-correlation (Moran's I) of regression residuals.

Behavior	Q_2 **of RSS**	Q_2 **of residuals**	z**-score**	p**-value**
Weak AP	-83	4	0.920	0.357
Strong AP	-74	6	7.702	≈ 0

Given the moderate correlation obtained in some of the analyses, and that the experiments were only performed in two environments, a reference point position determination method is not proposed. However, such determination method may have the following steps:

1) Place some reference points in the boundaries.
2) Distribute the rest of point maximizing the mean minimum distance among reference points.
3) Adjust the distribution to have some points closer to nearby APs.

4) Tend to LOS situations, assuring to place points in LOS and NLOS situations.

This work recommends the previous method steps as a set of guidelines that follow after the results of the analyses provided in this section. The most common approach of placing the reference points on a grid does not take into account the environment characteristics. The guidelines suggest adapting the sampling positions to the environment and highlight the importance of knowing the position of nearby antennas. Thus, the following two experiments address the environment aware regression and its evaluation on the Library environment. We selected the Library as the evaluation environment because the benefits from including environment knowledge into a regression model were expected to be greater for the Library than for Mannheim, as suggested by the correlations shown in Table III. Furthermore, the Library environment represents a medium-size open area with many obstacles (bookshelves), in which a positioning service is commonly desired.

D. Environment Aware regression assessment

The regression models were generated using the reference points that defined the boundary of the collection area (see Figure 2), which represent less than 8% of all available reference points. The remaining reference points were used to compute the regression residuals. The experiments only included APs that had measurements for all reference points.

Figure 5 presents the regression estimates for APs 15, 49 and 8 provided by a baseline that combines Natural Neighbour interpolation and Gradient Extrapolation and by the proposed regression model based on Support Vector Machine. For our proposed Model, the images were smoothed using 9 pixels square windows convolution.

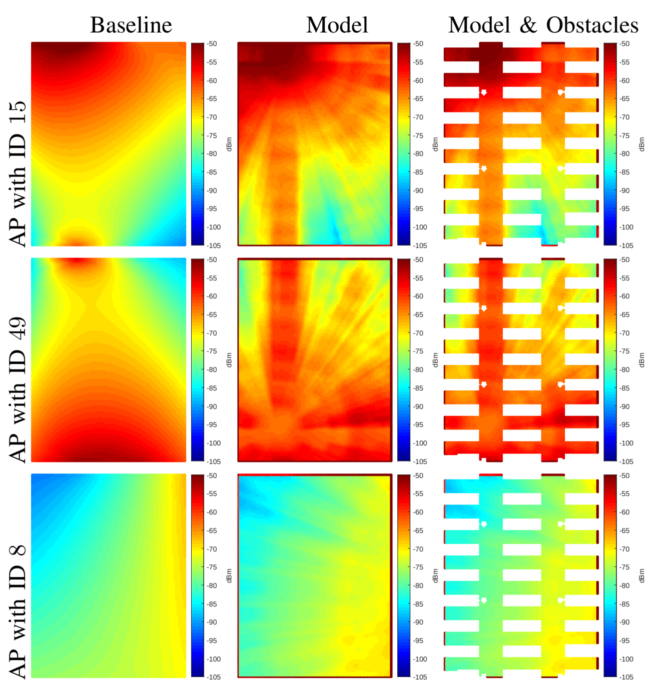
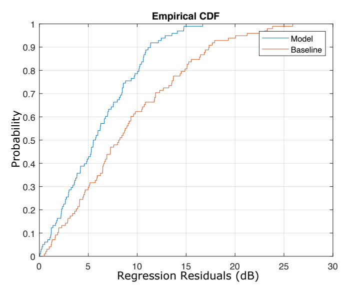


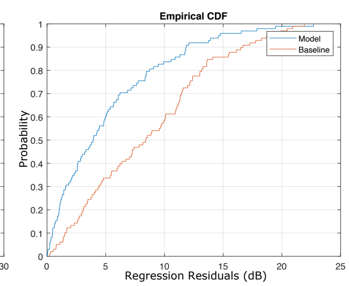
Fig. 5: Regression estimates for APs 15, 49 and 8.

Given the small number of training points, the two regressions performed remarkably well for the APs located inside the collection area, APs 15 and 49. The proposed regression can clearly capture the influence of obstacles in the radio map. For an AP outside the collection zone, the difference between Baseline and Model regression is not significant as the environment has little impact on the propagation of weak signals. The proposed model captures such behavior, and thus its estimates mostly depend on the distance to APs.

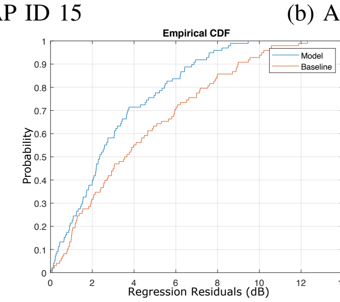
Figure 6 presents the regression residuals obtained using the baseline and our proposed model. The residuals obtained for the proposed model are consistently better than those from the baseline. For AP 15, the maximum residual value was about 10 dBm smaller in the proposed model than in the baseline. For AP 49, the maximum residual values were similar for the two approaches. However, the proposed model performed notably better than the baseline regarding percentiles between the 25^{th} and 75^{th}. For AP 8, the difference in residual values is less notable than for the previous two AP, which is in part a result of notably lower residual values.



(a) AP ID 15



(b) AP ID 49



(c) AP ID 8

Fig. 6: Regression residuals CDF of baseline and our model.

TABLE V: 75^{th} percentile of regression residuals in dB.

AP ID	1	6	8	15	17	49	51	52	54	69
Model	6.7	6.8	4.7	8.8	8.0	7.8	9.5	8.4	8.5	11.5
Baseline	10.7	10.3	6.6	13.6	13.2	12.2	13.2	10.1	11.2	16.2
Difference	4	3.5	1.9	4.8	5.2	4.4	3.7	1.7	2.7	4.7

Table V presents the 75^{th} percentile of regression residuals for the proposed model and the baseline method. The results are provided for some relevant APs, i.e., those APs with valid measurements available for all (106) reference points. Additionally, we included AP 71 (which had measurements for 105 points) and one weakly seen AP (AP 8). The proposed method performs better than the baseline for all selected APs.

V. Empirical validation

This section includes the empirical validation by applying together the two main contributions of this paper: the convenient positions where to collect the reference samples and the improved RSS regressor to enhance the radio map. For that purpose, we have used the data collected for the first month from the Library dataset [42]. It corresponds to a real environment with several obstacles (bookshelves and people) whose data collection was independent to this research work.
Traditional: The set 01 from the training set was used as reference data (radio map), and the sets 02–05 from the evaluation set were used for evaluation.
Measurement: Only the testing data (sets 01–05) was used for reference and evaluation. The 8 points highlighted in Figure 2 are used as training data (radio map), whereas the remaining points are used for evaluation.
Interpolation – Baseline: Similar to *Measurement*, but Natural Neighbors interpolation model is applied to increase the density of data in the training set.
Interpolation – Proposed model: Similar to *Measurement*, but our proposed interpolation model is applied to increase the density of data in the training set.

Following the ISO 18305 Standard for test and evaluation of localization and tracking systems, we report the results using the mean, median and 95th percentile (P95) of the positioning error in Table VI. Additionally, we provide the Third quartile (Q3) as done in the IPIN Competition [62] and the 90th percentile (P90).

TABLE VI: Results of the empirical evaluation

Base model.	Mean	Median	Q3	P90	P95
Traditional	3.41	2.83	4.74	6.63	7.94
Measurement	4.26	3.71	5.67	7.98	8.43
Interpolation – Baseline	4.06	3.69	5.67	7.32	8.7
Interpolation – Proposed model	3.94	3.8	5.38	6.82	7.21

As expected, the *traditional* approach, where multiple reference positions (24 in this case) are equally distributed in the operational area, is providing the best overall results, except, surprisingly, for the P95 metric. The *measurement* approach (with 8 reference points) is, as expected, providing the worst results as a few reference points are located in the periphery. Both interpolations, the *Natural Neighbors* and our proposed model, improve the results of the *measurement* approach. In general, our model is providing the best results using the reduced set of reference points. With a few reference points, we achieved a mean accuracy below 4 m and percentile errors close to the traditional approach.

Analysing the CDF plot (Figure 7) we can observe that: i) below 30th percentile, the traditional approach and both interpolations perform similarly; ii) between 30th and 80th percentiles, the traditional approach is clearly the best method (at the expense of collecting 3 times more reference data); and iii) the traditional approach and our proposed method have a similar performance in values above 80th percentile.

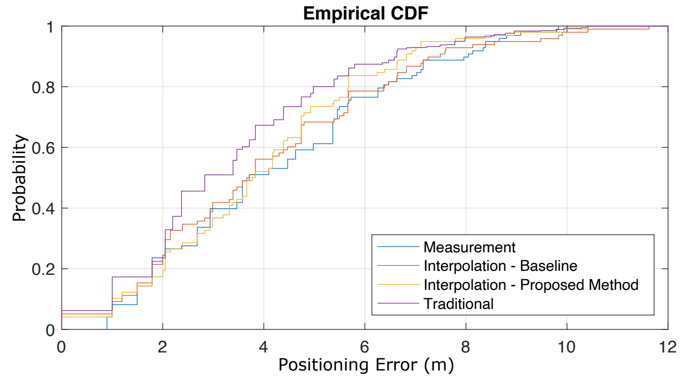


Fig. 7: Positioning accuracy.

VI. Conclusions

This paper has addressed the reduction of collection efforts for WiFi fingerprinting with two proposals. The first proposal is a set of guidelines to determine convenient positions where to collect WiFi samples. The second proposal is a model that improves the RSS regression estimates for APs that are strongly seen in the collection area. The guidelines were drawn from experiments that analyzed the effect that the distribution of collection points and the intensity of the APs in the environment have in (1) the accuracy of an IPS and (2) in the quality of a regression that could be applied to enrich the radio map. The guidelines highlight the importance of situating collection points around the boundaries of the target environment. Also, zones that are close to APs require more collection points than others. Thus, the position of an AP was shown to be an important piece of information for the determination of collection positions. Furthermore, the regressions and interpolation methods are shown to provide very good estimates for AP weakly seen in the environment.

The proposed model considers the influence of obstacles to improve WiFi RSS regressions for APs strongly seen in the environment. The model requires an approximate reference position of the AP whose RSS are to be estimated. The reference AP position and raw map information of the obstacles in the environment are used to create the training features for a Support Vector Machine regression. The regression proposal provided RSS estimates better than other regression or interpolation methods in the test environment and selected (strong) APs. The benefits of the regression proposal were also tested according to the positioning accuracy of a kNN method. The kNN was applied (1) using the radio map composed only by collected samples, (2) using the radio map created with other regression or interpolation methods, and (3) using the radio map created with our regression proposal. The best positioning accuracy was obtained using the third option.

The regression model presented in this paper could be considered a first step towards the definition of more general regression models or methods where, for instance, the type of material could be considered. To the best of this work's knowledge, there is no interpolation method, regression method, or tool that allows the direct modeling of the environment influence (presence of obstacles and walls) on a measured phenomenon.

The idea behind the regression model proposed in this paper could inspire others to include the environment characteristics into the existent methods that consider the spatial relation between measurements. We acknowledge that more ambitious conclusions would have reached with a more comprehensive evaluation. However, some methods proposed in the literature are not fully reproducible (some parameters are still missing) and the set of diverse data sets available for positioning do not contain enough information to integrate maps. We, the indoor positioning community, need to adopt and promote reproducible practices as well as creating rich data sets following international standards and ensuring interoperability. Further research is still needed to test the proposed method in a more challenging industrial environments and/or using BLE as positioning technology.

References

[1] G. M. Mendoza-Silva, J. Torres-Sospedra, and J. Huerta, "A meta-review of indoor positioning systems," *Sensors*, vol. 19, no. 20, 2019.

[2] R. Mautz, "Indoor positioning technologies," en, PhD thesis, ETH Zurich, Zurich, 2012.

[3] P. Bahl and V. Padmanabhan, "Radar: An in-building rf-based user location and tracking system," in *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications.*, IEEE, 2000.

[4] A. Pérez-Navarro, J. Torres-Sospedra, R. Montoliu, *et al.*, "Challenges of fingerprinting in indoor positioning and navigation," in *Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation*, Elsevier, 2019, pp. 1–20.

[5] E. S. Lohan, J. Torres-Sospedra, H. Leppäkoski, *et al.*, "Wi-fi crowdsourced fingerprinting dataset for indoor positioning," *Data*, vol. 2, no. 4, 2017.

[6] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Structured group sparsity: A novel indoor wlan localization, outlier detection, and radio map interpolation scheme," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 7, pp. 6498–6510, 2017.

[7] ——, "Modern wlan fingerprinting indoor positioning methods and deployment challenges," *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1974–2002, 2017.

[8] X. Zhou, T. Chen, D. Guo, *et al.*, "From one to crowd: a survey on crowdsourcing-based wireless indoor localization," *Frontiers of Computer Science*, vol. 12, no. 3, pp. 423–450, 2018.

[9] D. Zou, W. Meng, S. Han, *et al.*, "User aided self-growing approach on radio map construction for wlan based localization," in *Proceedings of the 26th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2013)*, 2013, pp. 991–997.

[10] S. Tsruya, R. Dalla Torre, D. Aljadeff, *et al.*, *Devices, methods, and systems for radio map generation*, US Patent 8,938,255, 2015.

[11] F. Meneses, A. Moreira, A. Costa, *et al.*, "4 - radio maps for fingerprinting in indoor positioning," in *Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation*, Academic Press, 2019, pp. 69 –95.

[12] Y. Zhuang, Z. Syed, J. Georgy, *et al.*, "Autonomous smartphone-based wifi positioning system by using access points localization and crowdsourcing," *Pervasive and Mobile Computing*, vol. 18, pp. 118 –136, 2015.

[13] J. Krumm and J. Platt, "Minimizing calibration effort for an indoor 802.11 device location measurement system," *Microsoft Research, November*, 2003.

[14] B. Li, Y. Wang, H. K. Lee, *et al.*, "Method for yielding a database of location fingerprints in wlan," *IEE Proceedings - Communications*, vol. 152, no. 5, pp. 580–586, 2005.

[15] Z. Gu, Z. Chen, Y. Zhang, *et al.*, "Reducing fingerprint collection for indoor localization," *Computer Communications*, vol. 83, pp. 56 –63, 2016.

[16] S. Ezpeleta, J. Claver, J. Pérez-Solano, *et al.*, "RF-Based Location Using Interpolation Functions to Reduce Fingerprint Mapping," en, *Sensors*, vol. 15, no. 10, pp. 27 322–27 340, 2015.

[17] J. Racko, J. Machaj, and P. Brida, "Wi-fi fingerprint radio map creation by using interpolation," *Procedia Engineering*, vol. 192, pp. 753 –758, 2017, 12th international scientific conference of young scientists on sustainable, modern and safe transport.

[18] M. Zhang and W. Cai, "Multivariate polynomial interpolation based indoor fingerprinting localization using bluetooth," *IEEE Sensors Letters*, vol. 2, no. 4, pp. 1–4, 2018.

[19] L. Xie, X. Jin, M. Zhou, *et al.*, "Cost-efficient ble fingerprint database construction approach via multi-quadric rbf interpolation," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, 2019.

[20] J Talvitie, M Renfors, and E. S. Lohan, "Distance-Based Interpolation and Extrapolation Methods for RSS-Based Localization With Indoor Wireless Signals," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1340–1353, 2015.

[21] V. Moghtadaiee, S. A. Ghorashi, and M. Ghavami, "New reconstructed database for cost reduction in indoor fingerprinting localization," *IEEE Access*, vol. 7, pp. 104 462–104 477, 2019.

[22] J. S. Seybold, *Indoor Propagation Modeling*. Wiley, 2005, ch. 9, pp. 208–216.

[23] M. U. Ali, S. Hur, and Y. Park, "Locali: Calibration-free systematic localization approach for indoor positioning," *Sensors*, vol. 17, 2017.

[24] A. S. Glassner, *An introduction to ray tracing*. Elsevier, 1989.

[25] M. Ayadi, N. Torjemen, and S. Tabbane, "Two-dimensional deterministic propagation models approach and comparison with calibrated empirical models," *IEEE Transactions on Wireless Communications*, vol. 14, no. 10, pp. 5714–5722, 2015.

[26] P. S. Heckhert, "Radiosity in flatland," *Computer Graphics Forum*, vol. 11, no. 3, pp. 181–192, 1992.

[27] M. F. Cohen, D. P. Greenberg, D. S. Immel, *et al.*, "An efficient radiosity approach for realistic image synthesis," *IEEE Computer graphics and Applications*, vol. 6, no. 3, pp. 26–35, 1986.

[28] Óscar Belmonte-Fernández, R. Montoliu, J. Torres-Sospedra, *et al.*, "A radiosity-based method to avoid calibration for indoor positioning systems," *Expert Systems with Applications*, vol. 105, pp. 89 –101, 2018.

[29] M. M. Atia, A. Noureldin, and M. J. Korenberg, "Dynamic online-calibrated radio maps for indoor positioning in wireless local area networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 9, pp. 1774–1787, 2013.

[30] P. Richter and M. Toledano-Ayala, "Revisiting gaussian process regression modeling for localization in wireless sensor networks," *Sensors*, vol. 15, no. 9, pp. 22 587–22 615, 2015.

[31] H. Zou, M. Jin, H. Jiang, *et al.*, "Winips: Wifi-based non-intrusive indoor positioning system with online radio map construction and adaptation," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8118–8130, 2017.

[32] W. Sun, M. Xue, H. Yu, *et al.*, "Augmentation of fingerprints for indoor wifi localization based on gaussian process regression," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 11, pp. 10 896–10 905, 2018.

[33] H. Ai, K. Tang, W. Huang, *et al.*, "Fast fingerprints construction via gpr of high spatial-temporal resolution with sparse rss sampling in indoor localization," *Computing*, 2019.

[34] C. Liu, A. Kiring, N. Salman, *et al.*, "A kriging algorithm for location fingerprinting based on received signal strength," in *2015 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 2015, pp. 1–6.

[35] S.-S. Jan, S.-J. Yeh, and Y.-W. Liu, "Received signal strength database interpolation by kriging for a wi-fi indoor positioning system," *Sensors*, vol. 15, no. 9, pp. 21 377–21 393, 2015.

[36] S. Kram, C. Nickel, J. Seitz, *et al.*, "Spatial interpolation of wi-fi rss fingerprints using model-based universal kriging," in *2017 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 2017, pp. 1–6.

[37] Y. Du, D. Yang, and C. Xiu, "A novel method for constructing a wifi positioning system with efficient manpower," *Sensors*, vol. 15, no. 4, pp. 8358–8381, 2015.

[38] N. Hernández, M. Ocaña, J. M. Alonso, *et al.*, "Continuous space estimation: Increasing wifi-based indoor localization resolution without increasing the site-survey effort," *Sensors*, vol. 17, no. 1, 2017.

[39] L. Li, G. Shen, C. Zhao, *et al.*, "Experiencing and handling the diversity in data density and environmental locality in an indoor positioning service," in *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, 2014, pp. 459–470.

[40] L. Kanaris, A. Kokkinis, G. Fortino, *et al.*, "Sample Size Determination Algorithm for fingerprint-based indoor localization systems," *Computer Networks*, vol. 101, pp. 169–177, 2016.

[41] W. Bong and Y. C. Kim, "Fingerprint wi-fi radio map interpolated by discontinuity preserving smoothing," in *International Conference on Hybrid Information Technology*, Springer, 2012, pp. 138–145.

[42] G. M. Mendoza-Silva, P. Richter, J. Torres-Sospedra, *et al.*, "Long-term WiFi fingerprinting dataset for research on robust indoor positioning," *Data*, vol. 3, no. 1, 2018.

[43] T. King, S. Kopf, T. Haenselmann, *et al.*, *CRAWDAD dataset mannheim/compass (v. 2008-04-11)*, Downloaded from https : / / crawdad.org/mannheim/compass/20080411, 2008.

[44] J. Torres-Sospedra, P. Richter, G. Mendoza-Silva, *et al.*, "Characterising the alteration in the ap distribution with the rss distance and the position estimates," in *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2018, pp. 1–8.

[45] T. King, T. Haenselmann, and W. Effelsberg, "On-demand fingerprint selection for 802.11-based positioning systems," in *International Symposium on a World of Wireless, Mobile and Multimedia Networks*, 2008.

[46] G. M. Mendoza-Silva, J. Torres-Sospedra, J. Huerta, *et al.*, "Situation goodness method for weighted centroid-based wi-fi aps localization," in *Progress in Location-Based Services 2016*, Springer, 2017, pp. 27–47.

[47] J. E. Bresenham, "Algorithm for computer control of a digital plotter," *IBM Systems journal*, vol. 4, no. 1, pp. 25–30, 1965.

[48] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel, "On the shape of a set of points in the plane," *IEEE Transactions on information theory*, vol. 29, no. 4, pp. 551–559, 1983.

[49] N. Metropolis *et al.*, "The beginning of the monte carlo method," *Los Alamos Science*, vol. 15, no. 584, pp. 125–130, 1987.

[50] A. Crooks, "Agent-based modeling and geographical information systems," *Geocomputation: a practical primer. Los Angeles, London, New Delhi, Singapore, Washington DC: Sage*, pp. 63–77, 2015.

[51] R. Sibson, "A brief description of natural neighbour interpolation," *Interpreting multivariate data*, 1981.

[52] T. Yang, *Finite element structural analysis*. Prentice-Hall Englewood Cliffs, NJ, 1986, vol. 2.

[53] D. Watson, *Contouring: a guide to the analysis and display of spatial data*. Elsevier, 1992, vol. 10.

[54] D. Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proceedings of the 1968 23rd ACM national conference*, ACM, 1968, pp. 517–524.

[55] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Berlin, Heidelberg: Springer-Verlag, 1995.

[56] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT Press Cambridge, MA, 2006, vol. 2.

[57] J. A. Nelder and R. W. Wedderburn, "Generalized linear models," *Journal of the Royal Statistical Society: Series A (General)*, vol. 135, no. 3, pp. 370–384, 1972.

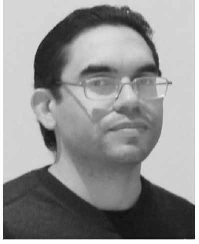
[58] L. Breiman, J. Friedman, C. Stone, *et al.*, *Classification and Regression Trees*, ser. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis, 1984.

[59] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[60] P. A. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, vol. 37, no. 1/2, pp. 17–23, 1950.

[61] A Goldsmith, "Path Loss and Shadowing," in *Wireless Communications*, Cambridge University Press, 2005, ch. 2, pp. 25–62.

[62] F. Potortì, A. Crivello, and F. Palumbo, "11 - the evaal evaluation framework and the ipin competitions," in *Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation*, ser. Intelligent Data-Centric Systems, J. Conesa, A. Pérez-Navarro, J. Torres-Sospedra, *et al.*, Eds., Academic Press, 2019, pp. 209 –224.



Germán Martín Mendoza-Silva has a Bachelor in Computer Science from the University of Oriente, Cuba in 2005; a Msc. in Geospatial Technologies from WWU (Germany), UNL (Portugal) and UJI (Spain) in 2015; and a PhD in Informatics from the UJI (Spain) in 2020. His research interests are on WLAN-based indoor positioning, indoor navigation, machine learning and GIS applications.

Ana Cristina Costa is Associate Professor at NOVA IMS, and Head of SIMAQ Quality Management System. She has a Ph.D. in Engineering Sciences (Technical University of Lisbon). The main scientific research area is spatial statistics, particularly the modelling of spatial-temporal phenomena using geostatistics. She has more than eighty refereed scientific publications, including a book, articles in international journals, and conference proceedings. Ana Costa has lectured over a dozen short courses in statistics topics, as well as courses on SAS System programming. She collaborated in several research and development projects, both national and international, and she coordinated the GSIMCLI research project funded by FCT (Portuguese Science Foundation).

Joaquín Torres-Sospedra received his PhD about Ensembles of Neural Networks and Machine Learning from Universitat Jaume I in 2011. Since January 2020 he is the Scientific Coordinator of UBIK Geospatial Solutions. He has authored more than 120 articles in journals and conferences. His current research interests include indoor positioning solutions based on Wi-Fi & BLE, Machine Learning and Evaluation. He is the chair of the IPIN International Standards Committee and IPIN off-site Competition.

Marco Painho is a Professor of Geographic Information Systems and Science at the Nova School of Information Management (NOVA IMS) of the Universidade Nova de Lisboa, Portugal. He holds a degree in Environmental Engineering by the Universidade Nova de Lisboa, a Master in Regional Planning by the University of Massachusetts, Amherst and a PhD in Geography by the University of California, Santa Barbara. He is the coordinator of the Master in Geographic Information Systems and Science and the Master of Science in Geospatial Technologies. He has over 30 years of experience in the GIS domain and coordinated over 100 projects in the application areas of the environment, natural resources management transportation, teaching among others. He is the author and editor of over 200 academic and professional publications.

Joaquín Huerta is a full professor at the Department of Information Systems from University Jaume I in Spain, where he teaches several courses related to GIS and Internet Technologies. His current research interests are indoor positioning, smart cities, mobile and web GIS applications and augmented reality. He is the head of the GEOTEC Research Group, Director of the Erasmus Mundus Master of Science in Geospatial Technologies degree program, run jointly with the universities of Münster and Nova de Lisboa. He is and has been the principal investigator of several research projects including EU projects as A-WEAR, GEO-C, EUROGEOSS. In addition to academic activities he is a founding member of UBIK Geospatial Solutions http://ubikgs.com