# Local-level Analysis of Positioning Errors in Wi-Fi Fingerprinting

Germán M. Mendoza-Silva*, Joaquín Torres-Sospedra† and Joaquín Huerta*
**Institute of New Imaging Technologies*, *Universitat Jaume I*, Castellón, España
†*UBIK Geospatial Solutions S.L.*, Castellón, España

***Abstract*—Nowadays, Location Based Services run over a net of heterogeneous devices (mainly smartphones) with different location capabilities thanks to, for instance, signals of opportunity as Wi-Fi. In contrast to professional deployments in controlled scenarios, the positioning error highly depends not only on the environment but also on the location. Traditional metrics for evaluating indoor positioning system may fail in obtaining lower-level details on the reported results. This paper introduces a way to perform a local-level analysis of the positioning errors. Our approach is based on analyses of the position-wise variance of positioning errors.**

## I. Introduction

In developed countries, people spend most of their time indoors [1], [2]. As the whole world develops, this tendency will expand to other countries. However, position estimation indoors still lacks a generally applicable, low-cost solution that allows location based services (LBS) tailored for indoor scenarios. The golden IPS should ideally be used in the way that GNSS (such as GPS) currently is used outdoors, but adapted to indoor requirements and challenges. It would be a technology, a technique, a method or their combination that feature a set of traits that can be inferred from the goal of obtaining the capabilities of GNSS for smartphones and from several common evaluation metrics for IPS [3].

Errors in position estimations are common to (indoor) positioning systems. Despite those errors may range from within only a few centimeters for technologies like UWB or Ultrasound, the most often used IPSs in wearable devices (e.g., smartphones) have error ranges of several meters.

The scenario where the positioning is performed may determine the significance of the magnitude of errors. The scenario usually influences the behavior of the IPS, as they usually rely on measurements of a signal –e.g., Wi-Fi or Bluetooth Low Energy in smartphone-based indoor positioning – that are heavily affected by the characteristics of the indoor environment. Moreover, the scenario is highly relevant given that, commonly, the position is determined for a person using a smartphone. Therefore, IPS evaluation is not only important for performing straightforward comparisons to determine which solution is the most adequate for a given application, but also an integral part of the development of an IPS. The proper evaluation of IPS proposals is one of the main challenges for indoor positioning, and the main subject of this work.

Previous works have shown that the positioning error might be extremely large when dealing with smartphone-based indoor positioning and Wi-Fi fingerprinting [4]. Despite Wi-Fi fingerprinting provides low general positioning error on the average case –within a few meters, as said before– it can reach values above 10 meters and, exceptionally, above 100 meters in very large testing environments [5].

The main metrics used to evaluate the indoor positioning systems –mainly the mean, median and percentiles– are usually a descriptor of the general positioning error. However, they are not able to tackle with smaller areas within the evaluation environment. e.g., a corner with low Wi-Fi coverage. This paper introduces a local-level analysis of positioning errors, which studies the variance of position estimates provided for the same test point and its behavior across an environment. The empirical data has come from the position estimates of the EvAAL-ETRI competition in 2015 [5] and the public UJIIndoorLoc dataset [6]. Having open-data allows the research community to rely on well-known datasets to assess new algorithms and perform comprehensive analysis.

The remainder of this paper is organized as follows. Section II describes the materials and methods used in this work. Section III introduces the local effects of positions estimations. Section IV discusses about the results on the study. Finally, Section V introduces the main conclusions of this work.

## II. Materials and Methods

To ensure the independence on the study of the local effects of position estimations, we rely on the publicly available UJIIndoorLoc database [6] and the position estimates of the EvAAL-ETRI competition in 2015 [5].

The UJIIndoorLoc database allows the evaluation of WiFi fingerprinting solutions on a complex indoor environment. The environment encompasses three buildings at UJI campus, each having 4-5 floors (see Figure 1). The positioning based on Wi-Fi fingerprinting is challenging. An access point (AP) seen with the strongest intensity is not necessarily the closest one as a result of intrinsic signal fluctuations and environment-induced variations, including absorption by people. Moreover, data was collected by crowd-sourced means by volunteers with different devices. The UJIIndoorLoc includes training (around 20.000 samples) and validation (around 1.000 samples).

Corresponding Author: J. Torres-Sospedra (torres@ubikgs.com)
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 Ciencia, Innovación y Universidades (INSIGNIA, PTQ2018-009981).

The UJIIndoorLoc database was used in the EvAAL-ETRI competition to evaluate off-line Wi-Fi fingerprinting solutions in 2015 [5]. For that purpose, they provided the competitors with a new blind test set with more than 5.000 unlabelled samples. The organizers of the competition allowed the competing teams to submit up to five sets of position estimates on the unlabelled test samples. Only the best set according to the mean accuracy was selected for competing in the final ranking.

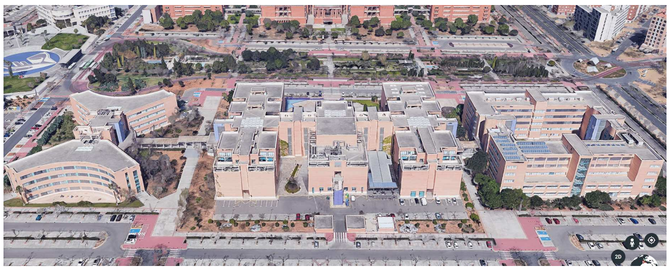

Fig. 1: 3D areal view, from Google Earth, of the three buildings considered in this section's experiments. From left to right, the buildings are identified as TI, TD and TC.

A total of four teams participated in the competition, namely "RTLSUM" [7], "HFTS" [8], "ICSL" [9], and "MOSAIC" [10]. The main statistics for the competition, mean positioning error and third quartile (75$^{\text{th}}$ percentile), are provided in Table I.

TABLE I: Team evaluation results, using best set of estimations according to the mean value of EvAAL error.

| Team | EvAAL Mean (m) | EvAAL 3rd quartile (m) |
|---|---|---|
| RTLSUM | 6.20 | 8.34 |
| ICSL | 7.67 | 10.87 |
| HFTS | 8.49 | 11.60 |
| MOSAIC | 11.64 | 10.65 |

Table I shows the competition results with an evaluation procedure based on the mean positioning error as in the EvAAL competition. The best set and the results use the mean positioning error as main evaluation metric, because it is a metric commonly used to evaluate indoor positioning systems as suggested in the *ISO/IEC 18305:2016* International Standard [11]. The table also shows the results using a evaluation procedure based on the third quartile (or 75$^{\text{th}}$ percentile) of the positioning error. In this later case, the best set and the results use the third quartile as it is the main metric used in the IPIN competition [12].

For both metrics, the positioning error was computed as the 2D Euclidean distance between the positions of the ground truth and estimated positions, with penalties added for floor error and building error of 4 m and 50 m, respectively. The positioning error was computed with the following equation:

$$error = \|\mathbf{P_R} - \mathbf{P_E}\| + p_f \cdot |f_R - f_E| + p_b \cdot (b_R! = b_E) \quad (1)$$

Where:

- $\mathbf{P_R}$ are the 2D coordinates of the evaluation point;
- $\mathbf{P_E}$ are the 2D coordinates estimated by the competitor;
- $\|\mathbf{P_R} - \mathbf{P_E}\|$ is the horizontal positioning errors and it is computed as the Euclidean distance between the location of the evaluation point and the estimated position;
- $p_f$ is the floor penalty – $p_f = 4$;
- $|f_R - f_E|$ is the absolute difference between the floor number of the evaluation point and the estimated floor number provided;
- $p_b$ is the building penalty – $p_b = 50$;
- $(b_R! = b_E)$ is 0 if the estimated building identifier matches the evaluation point, otherwise (in case of a wrong building identification) it is 1.

According to the results shown in Table I, the rankings provided by both metrics are not equivalent despite agreeing in the best method. The evaluation based on averaged values is more sensitive to extreme –either low or large– positioning errors. "MOSAIC" team provided locations in the wrong building for a very few cases, which largely penalized the averaged error. Thus, the selection of the accuracy metric for evaluation is very important. The local analysis selects the best set according to the mean error as done in the competition.

## III. Local Effects of Position Estimations

As shown in the previous section, the evaluation metric is of special relevance. One popular option for characterizing the general accuracy of an IPS is a CDF plot. CDF plots are very useful for determining the variability of the positioning errors of an IPS [13]. Also, they allow an easy determination of the number, or percentage, of position measurements that are below a given error. However, the CDF and related metrics like the mean, percentiles and ranges are broad descriptions for a large environment. It is difficult to directly translate that broad description to a smaller area given that IPS are commonly affected by the (local) environment characteristics.

The *ISO/IEC 18305:2016* International Standard [11] defines two metrics to measure IPS accuracy at a more local level than that provided by the CDF plot: the *circular error 95* and the *coverage*. The circular error 95 indicates the radius of the circle that, given a center position, contains 95% of position estimates. Instead of considering a whole environment, it can be applied individually to separate areas or individually to each test point if they contain several samples. The coverage metric indicates the ratio of evaluation points that meet a minimum performance requirement. The performance requirement is met for a test point having several samples if the maximum value of the error measurements for that point is below or equal to some non-negative value defined for the IPS.

This work addresses local analyses of errors measured for IPS. The IPS are those that participated in the EvAAL-ERTI competition in 2015. For the analyses, fingerprints and position estimates corresponding to some target reference points in the test set were selected. The targeted reference points are those where 5 or more fingerprints were collected. The targeted reference points accounted for 654 out of the 723 unique positions found in the private test set of the UJIIndoorLoc dataset. The circular error 95 was not computed given that the number of samples per reference point may be as low as 5 samples, i.e., a selection for 95% of estimates would mean a selection for the 100% of estimates. The computation of the coverage metric followed an adaptation from the definition of the *ISO/IEC 18305:2016* to consider the black box, off-line evaluation procedure followed in the competition.

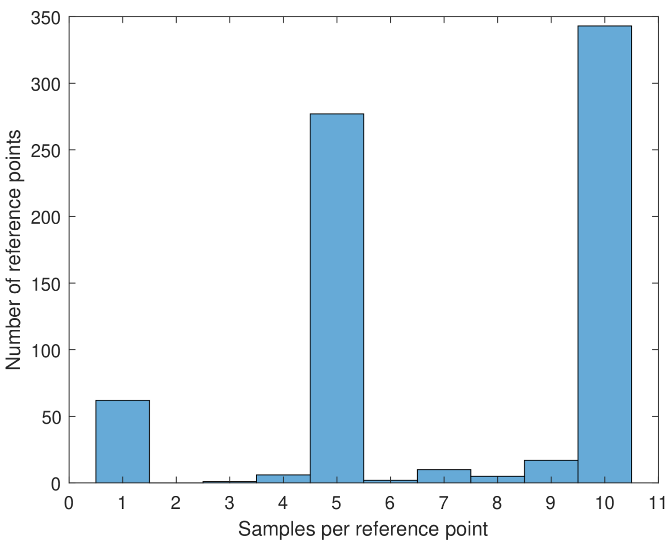


Fig. 2: Histogram of samples per reference point in the private test set of the UJIIndoorLoc dataset.

The number of reference points that have exactly five samples account for less than half of the total reference points, as presented in Figure 2. Thus, instead of using exactly five samples per point, five or more were used to increase the number of reference points available for the analysis and to avoid sampling randomness. Table II shows the results of computing the coverage metric for each team. For the minimum performance requirement, we choose the value 10 m, since it is the psychological barrier for a few meters.

TABLE II: Coverage metric adapted to the black box, off-line evaluation. The minimum performance requirement is 10 m.

| RTLSUM | ICSL | HFTS | MOSAIC |
|---|---|---|---|
| 0.62 | 0.50 | 0.58 | 0.46 |

In general, the coverage results indicate the presence of large errors (at least one case where the error is higher than 10 m) in about half of the reference points for most teams. The team RTLSUM is appreciatively the best performing according to coverage. The team HFTS performed better than the ICSL. The visualization of the coverage metric results can provide insights into the areas which insufficient IPS performance. Such visualization, despite it is recommended, is not provided here to maintain the privacy restrictions of the UJIIndoorLoc's evaluation set.

The coverage metric exposed that a notable number of reference points have large errors. Such fact is significant given that almost 75% of errors should be lower than the chosen value for the minimum performance requirement. Thus, uncertainty in positioning accuracy should exist for many reference points. The uncertainty in measurements is sometimes addressed using sensibility analysis. Wi-Fi fingerprinting models have a large number of input variables. Also, for large environments, some of the variables may strongly influence the position measurement in an area while being unimportant for other areas. Thus, the sensibility analysis should be done at a local scale, i.e., individually at small areas or each reference point. However, the black box, offline evaluation approach addressed in this section imposes hard constraints upon the IPS responses for the input variables. The addressed evaluation approach rules out the possibility of getting new position estimates for selected variations in the fingerprints.

Thus, instead of performing a sensibility analysis, the relation between the variation of error magnitudes and the variation of the RSS was studied. The variation of error magnitudes was computed as the standard deviation of errors at each reference point. The variation of the RSS (input signals) was computed as the maximum value of the distances among the fingerprints from each reference point. The chosen distance between two fingerprints was the Euclidean distance in the fingerprint space, which is a measure that is commonly applied as a similarity metric for Wi-Fi or BLE deterministic fingerprinting. The maximum fingerprint distance at each reference point ranged from 4.24 dBm to 163.8 dBm. The latter value is indicative of large variability in the input signals. The standard deviation of errors at each reference point ranged from $0$ to $224.57$ meters. The former value suggest a consistent estimation while the latter is indicative of very large and very low errors provided for the same reference point. Table III shows the $rho$ values of the Pearson correlation test applied to the previous two measures. The p-values were not shown as they were significant ($p < 0.05$) for the team RTLSUM and very significant ($p \ll 0.05$) for the other teams.

TABLE III: Pearson correlation $\rho$ values between standard deviation of errors and maximum fingerprint distance.

| RTLSUM | ICSL | HFTS | MOSAIC |
|---|---|---|---|
| 0.16 | 0.37 | 0.24 | 0.16 |

The correlation between the standard deviation of errors and maximum fingerprint distance is always positive, i.e., the larger the fingerprint distances are, the larger the error standard deviations are. Moreover, despite the correlation is weak for the RTLSUM and MOSAIC teams (both with 0.16), it is notable for the ICSL team (0.37). Although the ICSL team provided the lower floor detection rate in the competition [5], they applied deep learning models for feature selection. A large correlation may hint at a large susceptibility of a given IPS to variations in the input signals.

To further study the effect of variations of fingerprint distances on the position estimates, the standard deviation of error magnitudes was computed and later presented in Table IV for two cases. The first case was called “All RPs”. In this case, the standard deviation of error magnitudes of all reference points was considered and the median of those values is presented in the Table. We selected the median as it is less affected by outliers, i.e. extremely large positioning errors. The second case was called “Sel RPs”. In this case, only the values from selected references points are considered. The selection is performed attending to maximum fingerprint distance computed for each reference point. A reference point was selected if its maximum fingerprint distance was above a threshold value. The threshold was set as the median of the maximum fingerprint distances of all reference points. Thus, in the second case, we provide the the standard deviation of the reference points with the largest fingerprint variability. Finally, the relative difference between them is shown in the last row as their ratio to compare the different solutions provided by the competitors.

TABLE IV: Standard deviation of error magnitudes when considering all reference points (All RP) and when considering only those with large – above the median – standard deviation values of fingerprint distance (Sel RP).

| Selection | RTLSUM | ICSL | HFTS | MOSAIC |
|---|---|---|---|---|
| All RPs | 1.46 | 1.95 | 0.97 | 0.78 |
| Sel RPs | 1.67 | 2.43 | 1.19 | 1.87 |
| Ratio | 1.14 | 1.25 | 1.23 | 2.40 |

The values presented for the case “Sel RPs” are consistently larger than those for the case “All RPs”. The increase is particularly notable for the MOSAIC team, having increases larger than twice of those of the “All RPs” case. A likely reason for the MOSAIC team’s notable increase is that the selected reference points should be those where building miss-identification occurred, which are the ones that have the largest estimation errors. The second most notable increase corresponds to the ICSL team, which is the team that showed a larger correlation in Table III. The team with the least increase is RTLSUM, which had the smallest correlation in Table III and which was the one that provided best overall results across the error accuracy and coverage metrics.

## IV. Discussion

The results from Table III and Table IV suggest that the IPS from the RTLSUM team may be more robust to nominal changes in the input signals, i.e., changes related to short-term signal variations resulting from multi-path, device movement, or collection orientation, for example. Also, the results support a notion that is known for WiFi or BLE fingerprinting positioning: estimates corresponding to the same ground truth usually jump around its positions. Also, the results indicate that the instability of those jumps – random error and not bias – is related to distances among the operational fingerprints.

Reducing the magnitude of errors leads an IPS to perform better in any of the analyses explored in this section, as showed for the RTLSUM team. However, general metrics, like percentiles or their related CDF plot may hide local behaviors and are insufficient for analysing the robustness of a system. The *ISO/IEC 18305:2016* defines several conditions on the collection of test points for IPS evaluation, including for example the number or reference points. However, none of those recommendations can assure to cover the whole spectrum of nominal cases of signal readings, at least for fingerprinting-based systems. Therefore, further research efforts on IPS evaluation may dig into conditions on the evaluation data like, e.g., signal variations bounds applicable for a given environment and application. A dataset meeting such conditions may be used for evaluation instead of lots of other test data.

## V. Conclusions

This work addressed the evaluation of Indoor Positioning Systems. In particular, it presented a local-level analysis of positioning errors that studied the variance of position estimates provided for the same test point and its behavior across an environment. For that purpose, we have used a large public dataset and the position estimates provided by four teams that participated in the EvAAL-ETRI Competition in 2015. Therefore, raw data and the Indoor Positioning Systems are completely independent to the research performed. i.e. the research questions has not influence the way to calibrate and build up the different positioning solutions.

In general, the common metrics and CDF plots might hinder (including the average and median positioning error) some local behaviours as they summarize all positioning errors or visualize them without a connection with the location. The side-effect of this locality is that they might be insufficient for analysis the robustness of a particular indoor positioning system. As most of Wi-Fi fingerprinting solutions rely on a external wireless network infrastructure, locality is relevant to identify if the systems has a good general behavior as, in some scenarios, the reliability in all the operational area is a must.

Our recommendation is to ensure the quality of the evaluation data and, at the same time, stress the positioning solutions with, for instance, signal variations that mimic the real given environment. A comprehensive evaluation should not only consider the presence of corner cases (cases with low probability to happen but with large positioning errors) but also the procedures to detect them during the evaluation. The database stressed the evaluation by using a radio map built with more than 20 devices for the evaluation. Moreover, the geometry, the purpose and the APs distribution is completely heterogeneous in the three buildings represented in the dataset, so a single metric summarizing the positioning error does not represent the local characteristics across the operational area.

This work has opened a new research line on evaluation. As future work, we will work on defining additional metrics and procedures to have more detailed evaluation of the robustness and reliability of indoor positioning systems based on Wi-Fi (and, even, other positioning technologies) fingerprinting.

# REFERENCES


[1] N. E. Klepeis, W. C. Nelson, W. R. OTT, J. P. Robinson, A. M. Tsang, P. Switzer, J. V. Behar, S. C. Hern, and W. H. Eengelmann, "The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants," *Journal of Exposure Science & Environmental Epidemiology*, vol. 11, no. 3, pp. 231–252, Jul. 2001. [Online]. Available: https://doi.org/10.1038/sj.jea.7500165

[2] M. Śmiełowska, M. Marć, and B. Zabiegała, "Indoor air quality in public utility environments—a review," *Environmental Science and Pollution Research*, vol. 24, no. 12, pp. 11 166–11 176, 2017.

[3] A. Basiri, E. S. Lohan, T. Moore, A. Winstanley, P. Peltola, C. Hill, P. Amirian, and P. F. e Silva, "Indoor location based services challenges, requirements and usability of current solutions," *Computer Science Review*, vol. 24, pp. 1–12, 5 2017. [Online]. Available: https://doi.org/10.1016/j.cosrev.2017.03.002

[4] J. Torres-Sospedra and A. Moreira, "Analysis of sources of large positioning errors in deterministic fingerprinting," *Sensors*, vol. 17, no. 12, 2017. [Online]. Available: https://www.mdpi.com/1424-8220/17/12/2736

[5] F. Potortì, P. Barsocchi, M. Girolami, J. Torres-Sospedra, and R. Montoliu, "Evaluating indoor localization solutions in large environments through competitive benchmarking: The evaal-etri competition," in *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Oct 2015, pp. 1–10.

[6] J. Torres-Sospedra, R. Montoliu, A. Martinez-Uso, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems," in *International Conference on Indoor Positioning and Indoor Navigation*, 2014, pp. 261–270.

[7] A. Moreira, M. J. Nicolau, F. Meneses, and A. Costa, "Wi-fi fingerprinting in the real world - RTLS@UM at the EvAAL competition," in *Proceedings of the Sixth Conference on Indoor Positioning and Indoor Navigation*, 2015.

[8] S. Knauth, M. Storz, H. Dastageeri, A. Koukofikis, and N. Mähser-Hipp, "Fingerprint calibrated centroid and scalar product correlation rssi positioning in large environments," in *Proceedings of the Sixth International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2015.

[9] S. Choi, J. Yoo, and H. J. Kim, "Machine learning for indoor localization: Deep learning and semi-supervised learning," in *USB On-Site Proceedings of the Sixth International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2015.

[10] R. Berkvens, M. Weyn, and H. Peremans, "Localization Performance Quantification by Conditional Entropy," in *Proceedings of the Sixth International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2015.

[11] ISO Central Secretary, "Information technology – real time locating systems – test and evaluation of localization and tracking systems," International Organization for Standardization, Geneva, CH, Standard ISO/IEC 18305:2016, 2016. [Online]. Available: https://www.iso.org/standard/62090.html

[12] F. Potortì, A. Crivello, and F. Palumbo, "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. Academic Press, 2019, pp. 209 – 224. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780128131893000113

[13] F. Potortì, A. Crivello, P. Barsocchi, and F. Palumbo, "Evaluation of indoor localisation systems: Comments on the iso/iec 18305 standard," in *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 9 2018, pp. 1–7.