International Journal of Wireless and Ad Hoc Communication

♦ ISSN: 2692-4056 Vol: 11, No. 01, 2027 ♦ PP. 46–57

Article DOI: https://doi.org/10.54216/IJWAC.110105



FULL LENGTH ARTICLE

# Rank Fingerprints and Discounted Absence for Wi-Fi Positioning Across Devices and Access-Point Outages: Four Hypotheses Tested on a Multi-Building Dataset

**Amine Saddik**[1,*] **Ika Hesti Agustin**[2]

[1] Faculty of Applied Sciences, Ibn Zohr University, Ait Melloul, Morocco

[2] Department of Mathematics, University of Jember, Jember, East Java, Indonesia

**Emails:** EL.cherrat@gmail.com • ikahesti.0fmipa@unej.ac.id

Received: July 18, 2026 Revised: September 16, 2026 Accepted: November 09, 2026 ⋆ Corresponding author

## ABSTRACT

Wi-Fi fingerprinting is usually scored on the phones that built the radio map and with every access point alive. We state four hypotheses about what happens otherwise and test them on a public multi-building dataset, holding out one phone at a time and then scoring a campaign recorded months later. A nearest-neighbour locator is run on five fingerprint representations; a distance that discounts access points present in the map but absent from the query is introduced; and three ways of attaching an error radius to each fix are compared. Holding out a phone multiplies the mean error by more than four, most of it owed to unseen locations, while the loss in floor identification is owed mainly to the device. A rank-based fingerprint improves floor identification for twelve of sixteen phones and lowers the later-campaign error by a fifth, at a price when the phone is known. Offset correction does not help. Discounting absence keeps the error within a quarter of its clean value when three access points in ten are silent, where the standard distance more than doubles it. Neighbour spread alone is a poor guide to the error; a four-cue quantile model is better.

**Keywords:** Indoor positioning ▪ Wi-Fi fingerprinting ▪ Device heterogeneity ▪ Access-point outage ▪ Error estimation

## 1. INTRODUCTION

A Wi-Fi fingerprinting system answers a query by looking up the received signal strengths in a radio map recorded earlier [1,2]. Two assumptions sit underneath the lookup. The phone asking is taken to hear the access points as the survey phones heard them, and the access points heard during the survey are taken to be still there. Neither holds for long. Chipsets and antennas differ by several decibels and by the weakest signal they can report [3, 4], and access points are moved, replaced or switched off without the positioning service being told [5, 6].

Published accuracy figures rarely expose either problem. The common benchmark for multi-building positioning, UJIIndoorLoc [7], is most often scored by a random split of its training file or on its validation file alone, and a random split puts the same phone, the same reference point and often the same minute on both sides. The surveys of the field say as much [8, 9]. What is missing is an account that separates the sources of loss, tells a practitioner which cheap remedy addresses which source, and says how far a reported fix can be trusted once the conditions have moved.

This paper is organised around four hypotheses, fixed before the experiments were run and each stated so that the data could refute it.

**H1 (device gap).** Error measured with one phone held out exceeds error under random folds, beyond what the novelty

© 2027 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

of the held-out locations explains.

**H2 (invariant fingerprints).** A fingerprint built from the rank order of the access points narrows that gap relative to raw signal levels, whereas a fingerprint that only removes a per-phone offset does not.

**H3 (outage).** When a share of the access points is silent at query time, a distance that discounts map-only access points degrades more slowly than the Euclidean distance, at some cost when nothing is silent.

**H4 (self-assessment).** The spread of the nearest neighbours predicts the error of a fix well enough to give an error radius whose coverage holds on unseen phones.

The contribution is the test itself and three things that came out of it: a decomposition of the familiar optimism of random splits into a location part and a device part; a one-parameter distance, together with a bound that explains why rank fingerprints tolerate differences in receiver sensitivity; and a negative result on neighbour spread as an error estimate, with a remedy. H1 and H3 are supported. H2 holds in part, and H4 in its simple form is refuted. Section 2 gives the background, Section 3 the signal model and the method, Section 4 the experimental protocol, Section 5 the results hypothesis by hypothesis, Section 6 what a deployment can take from them, Section 7 the threats to validity, and Section 8 concludes.

## 2. BACKGROUND

### 2.1 Fingerprinting and its benchmark

RADAR [1] matched a query to the nearest stored fingerprints in signal space, and Horus [2] replaced the stored vectors by per-location distributions; nearest-neighbour matching [10] has nonetheless remained the reference method, and Torres-Sospedra et al. [11] showed how much its accuracy depends on the representation of the signal and on the distance. UJIIndoorLoc [7] made comparison possible at the scale of a campus: three buildings, four or five floors each, 520 access-point columns, a training campaign by many users and phones, and a later validation campaign recorded some three months afterwards. It was the basis of the 2015 EvAAL-ETRI competition [12]. Deep models followed, among them autoencoder classifiers [13], a scalable multi-label network [14], a convolutional model [15], recurrent and local variants [16, 17] and, recently, a heterogeneous graph network [18], each scored on the validation campaign and reporting an error between 7.8 m and 11.8 m there. We return to them in Section 5.8.

### 2.2 Device heterogeneity

Haeberlen et al. [3] observed that the readings of two cards are related almost linearly and calibrated one against the other. Calibration needs data from the new device. Calibration-free alternatives transform the fingerprint instead: Kjærgaard [4] used ratios between access points, Mahtab Hossain et al. [19] used pairwise signal-strength differences, which cancel an additive offset, and Machaj et al. [20] used ranks, which survive any monotone response. Zhang et al. [21] later matched fingerprints by grey relational analysis for the same purpose. Most of these proposals were validated with a handful of phones in one corridor or office. Whether they hold up over sixteen phones and three buildings, under a protocol that never shows the test phone to the map, is what H2 asks.

### 2.3 Outage and error estimates

Laoudias et al. [5] studied positioning with failed or corrupted access points and proposed detection and mitigation schemes; long-term datasets have since documented how much a deployment changes over months and years of service [6, 22]. On the side of self-assessment, Lemelson et al. [23] estimated the error of a fix from properties of the neighbour set, and conformal prediction [24–26] offers a way to turn any such score into a region with a stated coverage, as Nguyen [27] did for Wi-Fi and as Molina et al. [28] have done for RSSI-based indoor localization. Conformal guarantees assume exchangeability between calibration and test data [29]. A new phone months later is the case in which that assumption is in doubt, which is why H4 is tested there.

## 3. SIGNAL MODEL AND METHOD

### 3.1 Signal model

Let $s_j(\mathbf{p})$ be the power, in dBm, that an ideal receiver at position $\mathbf{p}$ would measure from access point $j$. Under the log-distance model [30],

$$s_j(\mathbf{p}) = P_j - 10 n_j \log_{10} \frac{\|\mathbf{p} - \mathbf{a}_j\|}{d_0} + \xi_j(\mathbf{p}), \tag{1}$$

with $P_j$ the power at reference distance $d_0$, $\mathbf{a}_j$ the position of the access point, $n_j$ a path-loss exponent and $\xi_j$ shadowing. A phone $d$ does not report $s_j$. It reports

$$r_j^{(d)} = g_d\big(s_j\big) + \varepsilon_j \quad \text{if } g_d(s_j) \geq \theta_d, \tag{2}$$

and nothing otherwise, where $g_d$ is a strictly increasing response, $\theta_d$ the sensitivity floor of the phone and $\varepsilon_j$ measurement noise. The set of access points reported is the detection set $\mathscr{D}$. Earlier work takes $g_d$ to be affine, $g_d(s) = \alpha_d s + \beta_d$ [3], and often assumes $\alpha_d = 1$ [19]; Equation (2) is more general and says two things that matter below, namely that devices differ in the values they report and in which access points they report at all.

### 3.2 Representations

A fingerprint is mapped to a vector $\boldsymbol{\phi} \in \mathbb{R}^A_{\geq 0}$ over the $A$ access points, with $\phi_j = 0$ for $j \notin \mathscr{D}$. Five maps are compared. The raw level rescales the reading,

$$\phi_j^{\text{raw}} = \frac{r_j + 105}{105}, \tag{3}$$

and the powed form of [11] is $\phi_j^{\text{pow}} = (\phi_j^{\text{raw}})^e$, which stresses strong signals. The relative level removes an additive offset by referring each reading to the mean $\bar{r}_{(3)}$ of the three strongest,

$$\phi_j^{\text{rel}} = \max\left(0.02,\ \frac{r_j - \bar{r}_{(3)} + 60}{60}\right). \tag{4}$$

The rank-decay form uses only the order. With $\rho_j \in \{0, 1, \dots\}$ the rank of access point $j$ among those detected, strongest first,

$$\phi_j^{\text{rank}} = \exp(-\rho_j / \tau), \qquad \tau = 5, \tag{5}$$

so that the weight of an access point halves about every three and a half ranks, and the one ranked twentieth carries about 2% of the weight of the strongest. The fifth map is detection only, $\phi_j = 1$ for $j \in \mathscr{D}$. The constant $\tau$ was fixed in advance, and its effect is reported in Section 5.3.

**Proposition 1.** *Let two phones observe the same position without noise, with responses $g_1, g_2$ strictly increasing. (i) If their detection sets coincide, their rank-decay fingerprints are identical. (ii) Their relative-level fingerprints, before clipping, coincide for all inputs if and only if $g_2 - g_1$ is constant. (iii) If the second phone detects only the $m$ strongest of the $M$ access points detected by the first, then*

$$\left\| \boldsymbol{\phi}_1^{\mathrm{rank}} - \boldsymbol{\phi}_2^{\mathrm{rank}} \right\|^2 = \sum_{i=m}^{M-1} e^{-2i/\tau} < \frac{e^{-2m/\tau}}{1 - e^{-2/\tau}}. \tag{6}$$

*Proof.* A strictly increasing map preserves order, so $\rho_j$ is the same under $g_1$ and $g_2$ whenever the same access points are ranked. This is (i). For (ii), $\phi^{\mathrm{rel}}$ depends on the readings only through differences $r_j - \bar{r}_{(3)}$; these are unchanged for every input exactly when $g_2(s) - g_1(s)$ does not depend on $s$. For (iii), a sensitivity floor removes the weakest access points and leaves the ranks of the others untouched, so the two vectors differ only in the removed coordinates, which hold $e^{-i/\tau}$ for $i = m, \dots, M-1$; the bound is the geometric tail. □

With $\tau = 5$ and $m = 10$ the right-hand side of Equation (6) is 0.056, against a squared norm of about 3.0 for a full fingerprint: a phone that hears only the ten strongest access points produces almost the fingerprint of a phone that hears thirty. The raw level has no such property, and neither has any representation under the opposite perturbation. If a strong access point falls silent, every weaker one moves up a rank and its coordinate grows by the factor $e^{1/\tau}$. That case is handled in the distance.

### 3.3 Distance with discounted absence

Let $\mathbf{q}$ be the query vector with detection set $\mathscr{D}_q$ and $\mathbf{x}$ a map vector with detection set $\mathscr{D}_x$. The squared Euclidean distance splits into three sums, over access points heard by both, by the query only and by the map only. We weight the third by $\lambda \in [0, 1]$:

$$d_\lambda^2(\mathbf{q}, \mathbf{x}) = \sum_{j \in \mathscr{D}_q \cap \mathscr{D}_x} (q_j - x_j)^2 + \sum_{j \in \mathscr{D}_q \setminus \mathscr{D}_x} q_j^2 + \lambda \sum_{j \in \mathscr{D}_x \setminus \mathscr{D}_q} x_j^2. \tag{7}$$

At $\lambda = 1$ this is the ordinary distance. At $\lambda = 0$ an access point that the map expects and the query lacks costs nothing. The asymmetry is deliberate. An access point heard by the query and unknown to the map at that place is evidence against the place, but an access point missing from the query has three possible causes (the query is elsewhere, the phone is less sensitive, or the access point is off), and only the first should count. With $\mathbf{m}_q$ the indicator vector of $\mathscr{D}_q$ and $\mathbf{x}^{\circ 2}$ the element-wise square, Equation (7) becomes

$$d_\lambda^2 = \|\mathbf{q}\|^2 - 2\,\mathbf{q}^\top \mathbf{x} + (1 - \lambda)\, \mathbf{m}_q^\top \mathbf{x}^{\circ 2} + \lambda\, \|\mathbf{x}\|^2, \tag{8}$$

which costs two matrix products against the map, where the Euclidean distance costs one. On the map used here a query takes 0.32 ms at $\lambda = 1$ and 0.63 ms otherwise, single-threaded.

The same quantities give a running estimate of how much of the map has gone silent. For a query, let $\mathscr{N}_0$ be its $k$ nearest entries at $\lambda = 0$, and define the miss statistic as the share of their squared norm carried by access points that the query does not hear at all,

$$u(\mathbf{q}) = \frac{1}{k} \sum_{i \in \mathscr{N}_0} \Big( 1 - \frac{\mathbf{m}_q^\top \mathbf{x}_i^{\circ 2}}{\|\mathbf{x}_i\|^2} \Big). \tag{9}$$

Its mean over a batch $\mathscr{B}$ of recent queries is mapped to an outage level through a line fitted on held-out phones with outage simulated at known levels,

$$\hat{q} = \frac{\bar{u}_{\mathscr{B}} - b_0}{b_1}, \tag{10}$$

and $\lambda$ is then read from a table of best discounts by level.

### 3.4 Position estimate and error radius

Let $\mathscr{N}$ be the $k$ map entries nearest to the query under $d_\lambda$, with positions $\mathbf{p}_i$ and weights $w_i \propto 1/(d_\lambda(\mathbf{q}, \mathbf{x}_i) + \varepsilon)$ normalised to one. The estimate and its neighbour spread are

$$\hat{\mathbf{p}} = \sum_{i \in \mathscr{N}} w_i\, \mathbf{p}_i, \qquad \sigma = \Big( \sum_{i \in \mathscr{N}} w_i\, \|\mathbf{p}_i - \hat{\mathbf{p}}\|^2 \Big)^{1/2}, \tag{11}$$

and building and floor are decided by weighted vote among the same neighbours. Three radii $R$ are compared. Each is meant to contain the true position with probability $1 - \alpha$, that is, the position should lie within $R$ of $\hat{\mathbf{p}}$ that often, and each is calibrated on errors $e_i = \|\hat{\mathbf{p}}_i - \mathbf{p}_i\|$ obtained with the phone of fingerprint $i$ held out. The constant radius is the empirical quantile, $R_{\mathrm{c}} = Q_{1-\alpha}\{e_i\}$. The spread-scaled radius follows the normalised conformal recipe [25],

$$R_\sigma = \hat{c}\,(\sigma + 1), \qquad \hat{c} = Q_{1-\alpha}\Big\{ \frac{e_i}{\sigma_i + 1} \Big\}, \tag{12}$$

with one metre added to keep the ratio finite. The third radius comes from a quantile regression [31] on four cues, $\mathbf{z} = (\sigma,\ d_{\min},\ |\mathscr{D}_q|,\ \max_j r_j)$, where $d_{\min}$ is the distance to the nearest neighbour. A gradient-boosted model [32] $f$ minimises the pinball loss

$$\ell_{1-\alpha}(e, f) = \max\big( (1 - \alpha)(e - f),\ \alpha\,(f - e) \big), \tag{13}$$

and its output is rescaled by a conformal factor computed from predictions made with each phone held out in turn, so that no phone calibrates its own radius,

$$R_{\mathrm{q}} = \hat{\kappa}\, f(\mathbf{z}), \qquad \hat{\kappa} = Q_{1-\alpha}\Big\{ \frac{e_i}{f_{-d(i)}(\mathbf{z}_i)} \Big\}. \tag{14}$$

Algorithm 1 summarises the pipeline.

**Algorithm 1.** Locating a query with rank fingerprints, discounted absence and an error radius

1. **Input:** $\{r_j\}_{j \in \mathscr{D}_q}$; map entries $(\mathbf{x}_i, \mathbf{p}_i, \ell_i)$; $k$, $\lambda$, $\tau$, $f$, $\hat{\kappa}$.
2. Rank the detected access points and form $\mathbf{q}$ by Equation (5).
3. Compute $d_\lambda^2(\mathbf{q}, \mathbf{x}_i)$ for all $i$ by Equation (8).
4. Select the $k$ nearest entries and their weights $w_i$.
5. Return $\hat{\mathbf{p}}$, $\sigma$ by Equation (11); building/floor by weighted vote.
6. Form $\mathbf{z}$ and return the radius $R_{\mathrm{q}}$ by Equation (14).

## 4. EXPERIMENTAL PROTOCOL

### 4.1 Data

UJIIndoorLoc [7] was taken from two public mirrors of the original files, which are identical byte for byte. Table 1 summarises it. The training campaign was recorded between 30 May and 20 June 2013 by 18 users with 16 phone models at 933 reference locations, and the validation campaign between 19 September and 8 October 2013 with 11 phone models, of which two also appear in training. We call the second the later campaign. Of the 520 access-point columns, 465 are detected at least once in training and form the feature space. The other 55 appear only in the later campaign, are unknown to the map and are ignored. Seventy-six training fingerprints detect no access point and were removed. Coordinates are in metres.

**Table 1.** The two campaigns of the dataset after cleaning

| | Training | Later campaign |
|---|---|---|
| Fingerprints | 19,861 | 1,111 |
| building 0 | 5,248 | 536 |
| building 1 | 5,159 | 307 |
| building 2 | 9,454 | 268 |
| Phone models | 16 | 11 |
| Reference locations | 933 | – |
| visited by two or more phones | 727 | – |
| Access points per fingerprint | 18.1 | 15.6 |
| Columns never detected (of 520) | 55 | 153 |

Two features of the collection shape what can be concluded. Phones and places are entangled: building 0 was surveyed by two phones only, and several phones worked on a single floor. And the later campaign already contains an outage of its own, for of the 465 access points of the training campaign 153 are never heard again, access points that account for 9.5% of all training detections.

### 4.2 Protocols

Four protocols are applied to every configuration. Random folds split the training fingerprints five ways at random. Location-grouped folds keep all fingerprints of a reference location in one fold, so that the test locations are new and the phones are not. Leave-one-device-out (LODO) holds out each of the 16 phones in turn. The later campaign is scored once, with the whole training set as the map. The number of neighbours was chosen per representation from $\{1,3,5,9,15\}$ by the LODO result and then frozen (3 for the raw and relative levels, 1 for the powed form, 5 for rank decay and detection only), which makes the LODO figures slightly favourable and leaves the later campaign untouched by any choice. To check that conclusions are not peculiar to nearest neighbours, the raw and rank representations were also fed to extremely randomised trees [33] of 100 estimators [34]. No setting was tuned on the later campaign.

Outage is simulated on the later campaign by silencing a random subset of the 465 access points in the queries only, each independently with probability $q \in \{0.1,\dots,0.5\}$, 20 draws per level. A targeted variant silences the $m$ strongest access points of each query, and as a reference we also report an informed system, one that knows which access points are silent and deletes them from the map before matching.

### 4.3 Scores and tests

A fix is scored by its horizontal error $e = \|\hat{\mathbf{p}} - \mathbf{p}\|$ in metres, by the building hit and by the floor hit, a floor counting only when the building is right. The combined score of the 2015 competition [12] adds penalties for the two kinds of miss,

$$E = e + 50 \cdot \mathbf{1}[\text{building wrong}] + 4 \cdot \mathbf{1}[\text{floor wrong}], \quad (15)$$

in metres. For radii we report the coverage $\Pr(e \le R)$, the mean radius and the mean pinball loss of Equation (13).

Representations are compared across phones with the Wilcoxon signed-rank test [35,36] on the 16 per-phone means, the phone being the unit that H2 is about; pooled means would let the two phones that hold 47% of the fingerprints decide. On the later campaign, intervals are percentile intervals over 2,000 bootstrap resamples of the queries [37], paired when two configurations are compared. Every test is two-sided. Grouped validation follows [38].

## 5. RESULTS

### 5.1 The device response

The model of Equation (2) can be checked directly at the 727 reference locations visited by more than one phone. For each phone, access point and location, the mean reading was compared with the mean over the other phones at that location (Figure 1); the median offsets so obtained run from $-6.9$ dB to $+6.6$ dB over the 15 phones that share locations with others. A straight-line fit of each phone's readings on the others' has a slope between 0.65 and 1.00 for 14 phones and of 0.26 for one. Slopes below one are biased downward by noise in the regressor, so their exact values should not be leaned on, but the spread among them is too wide for the unit-slope assumption behind offset correction. Sensitivity differs as well: the mean number of access points per fingerprint ranges from 10.2 for the least sensitive phone to 31.6 for the most (Table 4).

### 5.2 H1: the device gap

Table 2 and Figure 2 give the raw-level results under the four protocols. Random folds report a mean error of 2.21 m, a median of 0.29 m and a floor hit rate of 99.8%. The median says what is being measured: with some twenty fingerprints per reference location, the nearest neighbour of a held-out fingerprint is usually its sibling. With locations grouped, the mean error is 8.35 m and the floor rate 96.8%. With a phone held out, 10.07 m and 86.2%. On the later campaign, 8.85 m and 89.6%.

H1 is supported, and the decomposition is the more useful finding. Location comes first. Of the 7.86 m that separate random folds from LODO, 6.14 m appear as soon as the locations are new, with every phone still in the map. The device adds 1.72 m. For the floor the proportions reverse: 3.0 points are lost to new locations and a further 10.6 points to the new device. The same pattern holds when only the reference locations shared between phones are scored, 2.09 m

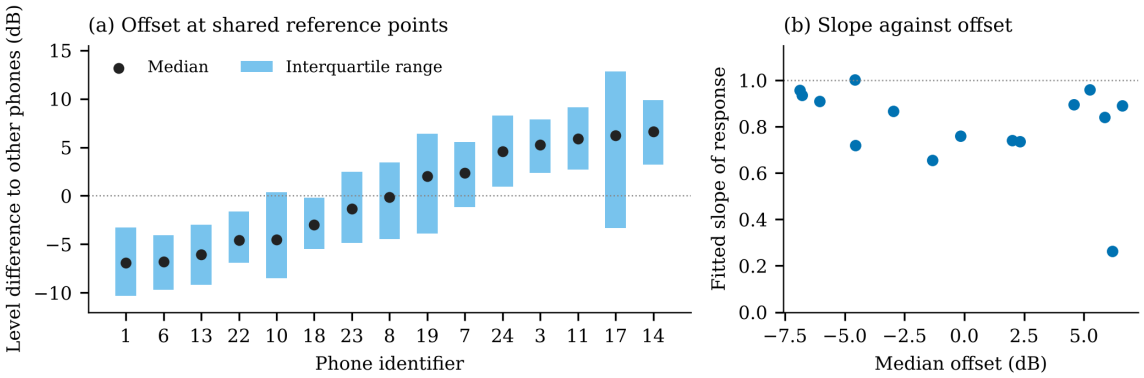



**Figure 1.** Per-phone difference in reported level against the other phones at shared reference locations (a), and the slope of each phone's fitted linear response against its median offset (b).

**Table 2.** Mean positioning error (m), floor hit rate (%) and combined score $E$ (m) of the nearest-neighbour locator by representation and protocol, at $\lambda = 1$. LODO: leave-one-device-out, pooled over fingerprints

| | Random folds | | | Location-grouped | | | LODO | | | Later campaign | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Representation** | **Error** | **Floor** | $E$ | **Error** | **Floor** | $E$ | **Error** | **Floor** | $E$ | **Error** | **Floor** | $E$ |
| Raw level | 2.21 | 99.8 | 2.21 | 8.35 | 96.8 | 8.64 | 10.07 | 86.2 | 11.10 | 8.85 | 89.6 | 9.39 |
| Powed | 2.25 | 99.7 | 2.27 | 10.01 | 96.9 | 10.26 | 12.92 | 86.2 | 14.39 | 8.77 | 92.9 | 9.18 |
| Relative level | 3.20 | 99.5 | 3.22 | 8.96 | 95.9 | 9.24 | 9.93 | 85.2 | 10.79 | 9.28 | 88.9 | 9.73 |
| Rank decay | 3.81 | 99.5 | 3.82 | 8.50 | 98.1 | 8.59 | 8.41 | 91.1 | 8.79 | 7.07 | 95.4 | 7.25 |
| Detection only | 4.39 | 97.3 | 4.52 | 10.19 | 90.0 | 10.67 | 11.72 | 76.5 | 12.89 | 12.55 | 76.2 | 13.91 |

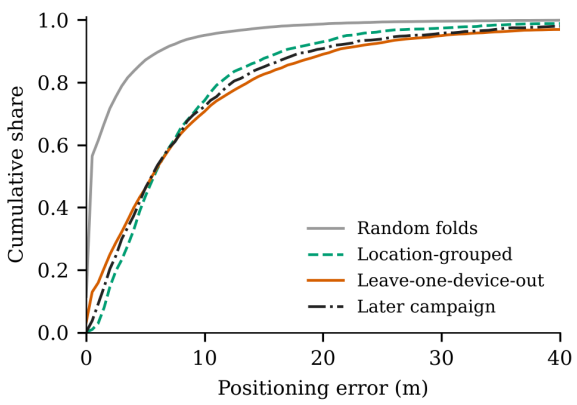



**Figure 2.** Cumulative distribution of the positioning error of the raw-level locator under the four protocols.

under random folds, 7.96 m with locations grouped and 9.06 m with the phone held out, so the device component is not produced by the held-out phone having visited places of its own. Those places exist, and they are costly. On the 12% of fingerprints at locations visited by a single phone, the LODO error is 17.5 m.

A random split therefore overstates accuracy mainly because it tests on replicas. What a new phone costs is the floor.

### 5.3 H2: invariant fingerprints

The rank-decay fingerprint behaves as H2 predicts in three of the four columns of Table 2 (Figure 3).

Under LODO it lowers the pooled mean error from 10.07 m to 8.41 m and raises the floor rate from 86.2% to 91.1%. On the later campaign the error falls from 8.85 m to 7.07 m, a paired difference of −1.78 m with a bootstrap interval from −2.52 to −1.12, and the floor rate rises by 5.8 points (interval 4.2 to 7.4). The gain is largest where the phone hears little. For the 118 later-campaign queries with eight access points or fewer, the mean error is 16.3 m with raw levels and 8.8 m with ranks, as part (iii) of Proposition 1 would lead one to expect. Under random folds the ranking inverts, 3.81 m against 2.21 m: when the phone is known, its absolute levels are information, and ranks throw it away.

Per phone, the picture is less tidy (Figure 4, Table 3). Ranks improve the floor rate for 12 of the 16 phones (Wilcoxon $p = 0.021$), but they improve the mean error for only 8 of the 16, and on that score the signed-rank test finds no difference between the two representations ($p = 0.30$). The pooled gain of 1.66 m comes largely from one heavily used phone, number 13, whose error halves from 13.1 m to 6.7 m; phones 7, 11 and 14, among others, do worse with ranks. Medians agree with this reading: the pooled LODO median is 5.58 m for raw levels and 5.85 m for ranks, and what ranks remove is the tail, the 95th percentile falling from 30.1 m to 24.2 m.

The second half of H2 holds without qualification. The relative level, which corrects an offset and nothing else, is no better than the raw level under LODO (9.93 m, 7 of 16 phones) and slightly worse on the later campaign, by 0.4 m. Given the slopes of Section 5.1 this is unsurprising. The powed form is significantly worse across phones (3 of 16, $p = 0.011$), although it does well on the later campaign's floors; amplifying strong signals amplifies the differences between receivers.

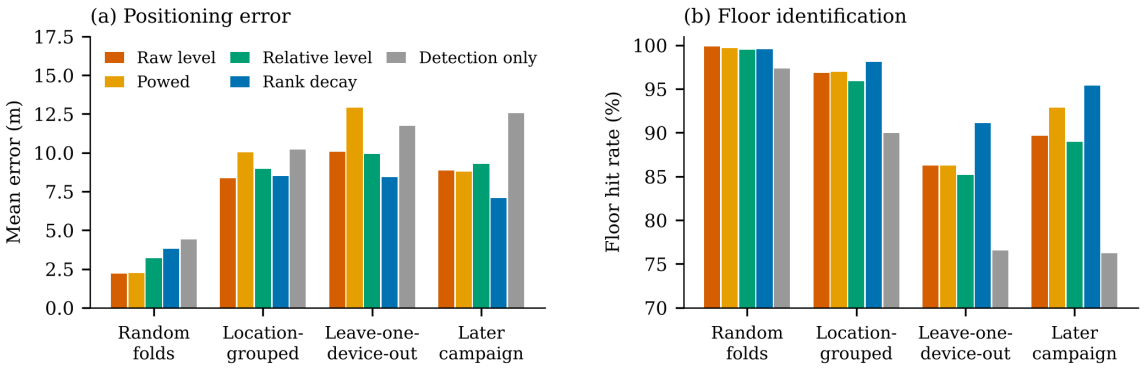



**Figure 3.** Mean positioning error (a) and floor hit rate (b) of the five representations under the four protocols.

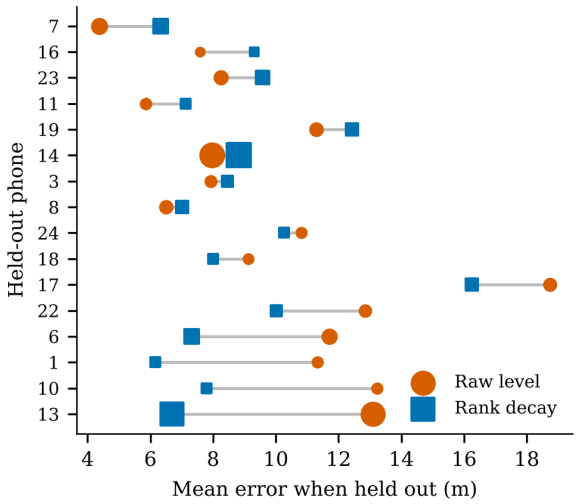



**Figure 4.** Mean error of each phone when held out, with raw-level and rank-decay fingerprints. Marker area grows with the number of fingerprints; phones are ordered by the change.

**Table 4.** Each phone held out in turn: fingerprints, access points heard per fingerprint (APs), mean error in metres and floor hit rate (%) with raw-level and rank-decay fingerprints

| | | | Error | | Floor | |
|---|---|---|---|---|---|---|
| Phone | $n$ | APs | Raw | Rank | Raw | Rank |
| 1 | 448 | 11.4 | 11.3 | 6.2 | 97.8 | 99.1 |
| 3 | 610 | 17.2 | 7.9 | 8.5 | 71.5 | 70.3 |
| 6 | 1,382 | 17.1 | 11.7 | 7.3 | 77.3 | 92.5 |
| 7 | 1,596 | 23.7 | 4.4 | 6.3 | 98.9 | 98.9 |
| 8 | 913 | 17.4 | 6.5 | 7.0 | 99.0 | 99.3 |
| 10 | 440 | 14.3 | 13.2 | 7.8 | 85.5 | 98.6 |
| 11 | 498 | 22.4 | 5.9 | 7.1 | 98.8 | 99.2 |
| 13 | 4,514 | 15.6 | 13.1 | 6.7 | 87.5 | 97.1 |
| 14 | 4,835 | 18.0 | 8.0 | 8.8 | 88.4 | 89.0 |
| 16 | 192 | 19.2 | 7.6 | 9.3 | 0.5 | 1.6 |
| 17 | 841 | 22.6 | 18.7 | 16.2 | 69.0 | 69.1 |
| 18 | 374 | 18.9 | 9.1 | 8.0 | 100.0 | 97.9 |
| 19 | 980 | 22.2 | 11.3 | 12.4 | 65.3 | 79.5 |
| 22 | 710 | 10.2 | 12.8 | 10.0 | 97.2 | 96.8 |
| 23 | 1,091 | 17.2 | 8.3 | 9.6 | 86.3 | 91.9 |
| 24 | 437 | 31.6 | 10.8 | 10.3 | 88.8 | 98.2 |

**Table 3.** Paired comparison with the raw level across the 16 held-out phones (Wilcoxon signed-rank), and on the later campaign

| Form | Error lower phones ($p$) | Floor higher phones ($p$) | Later campaign Δ error (m) estimate [95% CI] |
|---|---|---|---|
| Powed | 3 (0.011) | 11 (0.562) | $-0.08$ $[-0.79, 0.64]$ |
| Relative | 7 (0.597) | 6 (0.252) | $+0.43$ $[-0.16, 1.02]$ |
| Rank | 8 (0.298) | 12 (0.021) | $-1.78$ $[-2.52, -1.12]$ |
| Detection | 1 ($< 0.001$) | 0 ($< 0.001$) | $+3.70$ $[2.97, 4.50]$ |

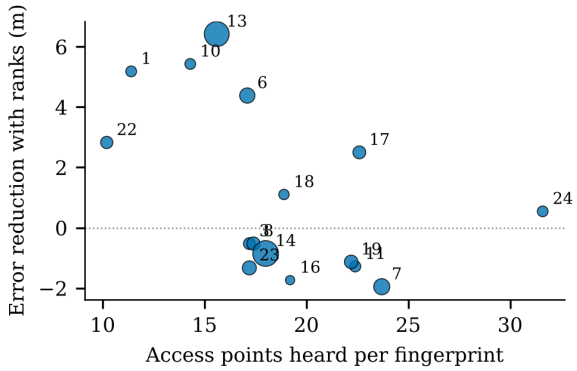



**Figure 5.** Reduction in mean error when ranks replace raw levels, for each held-out phone, against the mean number of access points that phone hears. Marker area grows with the number of fingerprints.

Table 4 lists the phones one by one, and a pattern can be read off it. The four phones that hear fewer than 16 access points per fingerprint (1, 10, 13 and 22) all gain from ranks, by 2.8 m to 6.4 m. Of the five that hear more than 20, three lose. A sensitive phone held out of the map is a phone whose weak detections the map cannot match, and raw levels handle that case tolerably; an insensitive phone is the case of Proposition 1(iii). Figure 5 plots the relation. Across the 16 phones the Spearman correlation between access points heard and metres gained is $-0.62$ ($p = 0.010$). The check was made after the fact, and it agrees with the bound.

On the later campaign the gain is likewise uneven across buildings: the mean error with raw levels and with ranks is

6.0 m and 5.8 m in building 0, 10.8 m and 8.6 m in building 1, and 12.4 m and 7.8 m in building 2, and the floor rate in building 1 rises from 75.6% to 91.2%.

One phone resists every representation. Phone 16 contributed 192 fingerprints, all on the top floor of one building, and when it is held out the floor is right once with raw levels and three times with ranks, out of 192. Two other phones surveyed that floor. Whatever the cause, it is outside what a change of fingerprint can repair, and it is a reminder that LODO on this dataset measures phone and place together.

Where the gain comes from can be read from the misses themselves (Table 5). With a phone held out, raw levels put 206 fingerprints in the wrong building and 2,525 on the wrong floor of the right building; ranks cut these to 11 and 1,752, a fall of 95% and 31%. More than nine wrong floors in ten are adjacent ones under either representation. The last two rows of the table are the telling ones. Among fixes on the right floor, the mean horizontal error is 7.69 m with raw levels and 7.66 m with ranks, whereas a wrong-floor fix carries about twice that. Ranks, then, do not sharpen positioning within a floor. They prevent fixes from landing on another floor or in another building, and the horizontal gain of Table 2 is a by-product.

**Table 5.** Anatomy of the misses with raw-level and rank-decay fingerprints

| | LODO | | Later campaign | |
|---|---|---|---|---|
| | Raw | Rank | Raw | Rank |
| Fixes | 19,861 | 19,861 | 1,111 | 1,111 |
| Wrong building | 206 | 11 | 3 | 0 |
| Wrong floor, right building | 2,525 | 1,752 | 112 | 51 |
| of which adjacent floor | 2,329 | 1,621 | 99 | 49 |
| Error, floor right (m) | 7.69 | 7.66 | 7.46 | 6.66 |
| Error, floor wrong (m) | 14.47 | 15.44 | 16.72 | 15.61 |

Two checks bear on generality. Extremely randomised trees show the same three-way pattern: ranks cost accuracy under random folds (3.60 m against 3.02 m), gain under LODO (9.75 m against 11.29 m, floor 91.3% against 86.1%) and gain on the later campaign (8.47 m against 9.89 m, floor 94.8% against 90.5%). And the decay constant matters within limits. With $\tau$ at 2, 3, 5, 8 and 12 the LODO error is 11.8, 10.0, 8.4, 7.8 and 8.0 m and the later-campaign error 8.5, 7.9, 7.1, 7.2 and 7.9 m, so that anything between 5 and 8 serves. The value fixed in advance was not the LODO optimum.

H2 is supported for floor identification and for the later campaign, and not established for horizontal error phone by phone.

### 5.4 An exploratory hybrid

The inversion under random folds and the uneven per-phone record suggest keeping both kinds of information. After the planned experiments we therefore tried one further fingerprint, the concatenation $[a\boldsymbol{\phi}^{\text{raw}}, \boldsymbol{\phi}^{\text{rank}}]$ with a weight $a$ on the raw part, at $k = 5$. This was not among the hypotheses, and three weights were tried. It is a lead, no more.

**Table 6.** Hybrid of raw-level and rank-decay fingerprints at $k = 5$: mean error (m), with phones improved over the raw level under LODO and the Wilcoxon $p$-value

| Fingerprint | Random folds error (m) | LODO error (m) | Later campaign error (m) |
|---|---|---|---|
| Raw level | 2.21 | 10.07 | 8.85 |
| Rank decay | 3.81 | 8.41 | 7.07 |
| Hybrid, $a = 0.5$ | 3.35 | 8.12 | 6.87 |
| Hybrid, $a = 1$ | 2.80 | 8.00 | 6.80 |
| Hybrid, $a = 2$ | 2.43 | 8.15 | 7.15 |

| Fingerprint | Error lower phones ($p$) | Floor higher phones ($p$) |
|---|---|---|
| Raw level | – | – |
| Rank decay | 8 (0.298) | 12 (0.021) |
| Hybrid, $a = 0.5$ | 8 (0.144) | 14 (0.003) |
| Hybrid, $a = 1$ | 10 (0.051) | 15 (0.001) |
| Hybrid, $a = 2$ | 14 ($< 0.001$) | 15 ($< 0.001$) |

Table 6 shows that the hybrid removes most of the price and most of the unevenness. At $a = 2$ it gives up 0.2 m to raw levels under random folds, where ranks alone give up 1.6 m, and with a phone held out it lowers the mean error for 14 of the 16 phones and raises the floor rate for 15. On the later campaign every weight is ahead of raw levels by 1.7 m to 2.1 m, and $a = 1$ at 6.80 m is the lowest error in this paper, though its floor rate of 93.4% is below the 95.4% of ranks alone. The weight trades the two: more raw level helps the known phone and the horizontal error, less of it helps the floor.

### 5.5 H3: outage and discounted absence

Figure 6 and Table 7 show what silent access points do. The standard configuration, raw levels with $\lambda = 1$, loses control quickly: 8.85 m with nothing silent, 14.5 m with one access point in five silent, 19.2 m at three in ten and 42.8 m at one in two, by which point the floor is right for 56% of queries and the building is sometimes wrong. Rank fingerprints with $\lambda = 1$ degrade more slowly, to 10.1 m at three in ten, since a missing access point shifts ranks where it would otherwise leave a hole of full depth.

Discounting absence changes the slope. With ranks and $\lambda = 0.25$ the error at three in ten silent is 8.62 m, 21% above the clean value, where the standard configuration is 117% above its own; the floor rate stays at 91.9%. The informed system, which is told what is off, reaches 8.05 m at that level, so a blind discount recovers about three quarters of what knowing the outage would give relative to ranks with $\lambda = 1$. The targeted outage is harsher. With the strongest access point of every query silent, ranks with $\lambda = 0.25$ lose 0.6 m and 1.4 floor points while raw levels with $\lambda = 1$ lose 5.0 m and 9.3 points; with the two strongest silent, every configuration is above 12.9 m and no floor rate exceeds 80%. Nearest-access-point failures in pairs are beyond what any of these distances absorbs.

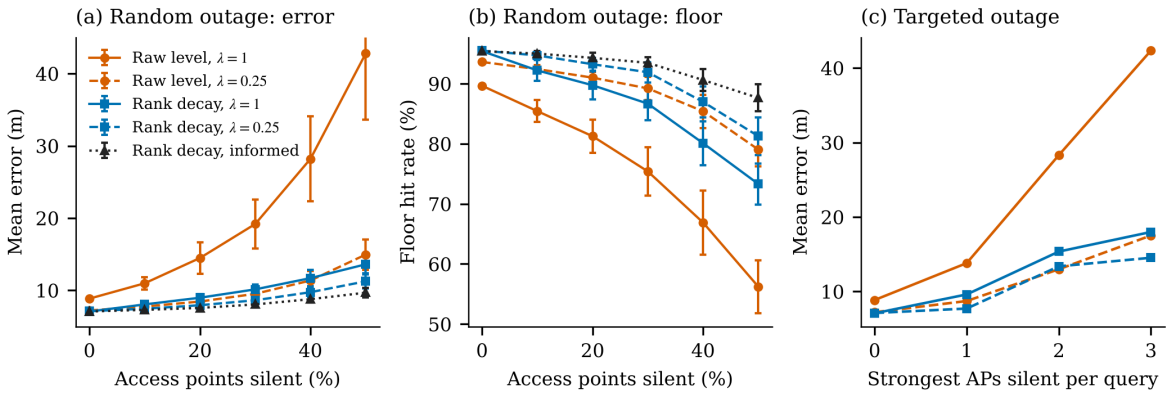



**Figure 6.** Later-campaign mean error (a) and floor hit rate (b) against the share of access points silenced at random, mean and standard deviation over 20 draws; mean error when the strongest access points of each query are silenced (c). Solid lines: $\lambda = 1$; dashed: $\lambda = 0.25$.

**Table 7.** Later campaign under random access-point outage: mean error in metres, with the floor hit rate (%) in parentheses, averaged over 20 draws

| | | Share of access points silent | | | | | |
|---|---|---|---|---|---|---|---|
| **Representation** | **Distance** | **0** | **10%** | **20%** | **30%** | **40%** | **50%** |
| Raw level | $\lambda = 1$ | 8.85 (89.6) | 10.98 (85.4) | 14.48 (81.3) | 19.18 (75.4) | 28.20 (66.9) | 42.79 (56.2) |
| Raw level | $\lambda = 0.25$ | 7.21 (93.6) | 7.76 (92.4) | 8.45 (91.0) | 9.52 (89.2) | 11.36 (85.4) | 14.95 (79.0) |
| Rank decay | $\lambda = 1$ | 7.07 (95.4) | 8.05 (92.2) | 8.98 (89.7) | 10.13 (86.7) | 11.69 (80.1) | 13.60 (73.3) |
| Rank decay | $\lambda = 0.25$ | 7.11 (95.5) | 7.45 (94.7) | 7.94 (93.2) | 8.62 (91.9) | 9.74 (87.0) | 11.25 (81.3) |
| Rank decay | $\lambda = 0$ | 7.33 (95.5) | 7.40 (95.1) | 7.71 (94.4) | 8.32 (93.4) | 9.30 (89.4) | 10.67 (84.4) |
| Rank decay | informed map | 7.07 (95.4) | 7.30 (95.0) | 7.57 (94.3) | 8.05 (93.5) | 8.77 (90.6) | 9.68 (87.6) |

The first column of Table 7 holds a result we had not expected. With no simulated outage at all, the discount improves the raw-level locator from 8.85 m to 7.21 m and its floor rate from 89.6% to 93.6%, nearly the whole gain that rank fingerprints provide on this campaign. The later campaign is not outage-free. A tenth of the training detections belong to access points that had disappeared by September, and the later phones hear fewer access points per fingerprint (Table 1). Much of what looks like device heterogeneity on this benchmark is therefore a difference in detection sets, and either remedy addresses it, ranks through Proposition 1(iii) and the discount by construction.

Is there a price? Table 8 and Figure 7 answer with the phone held out inside the training campaign, where no access point has been lost. For ranks there is: at no outage the error rises from 8.41 m at $\lambda = 1$ to 9.12 m at $\lambda = 0.25$ and 10.07 m at $\lambda = 0$, because absence is informative when the infrastructure is intact and every access point answers. The best $\lambda$ falls as outage grows, to 0.5 at one in ten silent, 0.25 at one in five and 0 at three in ten. For raw levels the discount pays even at no outage (10.07 m to 8.37 m), which again points at detection sets. H3 is supported in both of its clauses for ranks. The choice of $\lambda$ is thus an insurance decision: with ranks, $\lambda = 0.25$ costs 0.7 m when phones change and nothing fails, costs nothing measurable on the later campaign, and saves 1.5 m there when three access points in ten are silent.

**Table 8.** Mean error (m) with one phone held out and a share of access points silenced (three draws), by absence discount. Bold: best in row

| | **Silent** | $\lambda = 1$ | 0.5 | 0.25 | 0 |
|---|---|---|---|---|---|
| Rank decay | 0 | **8.41** | 8.68 | 9.12 | 10.07 |
| | 10% | 8.97 | **8.84** | 9.07 | 9.78 |
| | 20% | 10.78 | 9.88 | **9.63** | 9.81 |
| | 30% | 13.51 | 11.79 | 11.13 | **10.95** |
| Raw level | 0 | 10.07 | 8.91 | **8.37** | 11.24 |
| | 10% | 12.70 | 10.00 | **8.93** | 11.66 |
| | 20% | 16.97 | 12.06 | **10.45** | 12.40 |
| | 30% | 21.52 | 13.80 | **10.85** | 11.69 |

### 5.6 Setting the discount online

A deployment does not know its outage level, so the estimator of Equation (10) was tried as a way to set $\lambda$ without being told. The calibration line, fitted under LODO at simulated levels from 0 to 40%, has $b_0 = 0.129$ and $b_1 = 0.61$: with nothing silent, a held-out phone already misses 12% of the rank-weighted mass of its neighbours, and each ten points of outage add about six. The rule read from Table 8 is $\lambda = 1$ for $\hat{q} < 0.05$, 0.5 up to 0.15, 0.25 up to 0.25 and 0 beyond. Each batch is one draw of the later campaign.

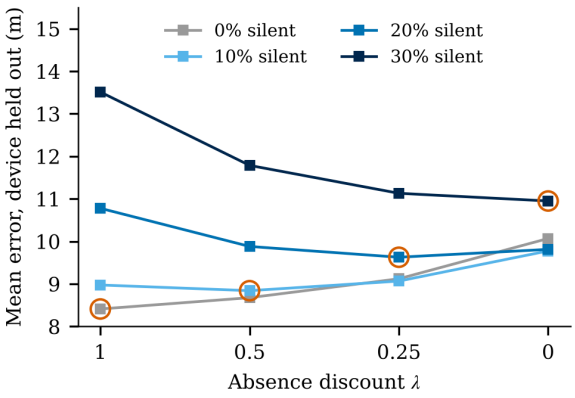



**Figure 7.** Leave-one-device-out error of the rank-decay locator against the absence discount, at four outage levels. Circles mark the best discount at each level.

**Table 9.** Online choice of the discount on the later campaign (rank decay): estimated outage level, and mean error in metres against two fixed discounts and the best fixed discount at each level

| True $q$ | $\hat{q}$ | Online | $\lambda = 1$ | $\lambda = 0.25$ | Best fixed |
|---|---|---|---|---|---|
| 0 | −0.08 | 7.07 | 7.07 | 7.11 | 7.07 |
| 0.10 | 0.06 | 7.73 | 8.05 | 7.45 | 7.40 |
| 0.20 | 0.15 | 8.14 | 8.98 | 7.94 | 7.71 |
| 0.30 | 0.25 | 8.47 | 10.13 | 8.62 | 8.32 |
| 0.40 | 0.36 | 9.30 | 11.69 | 9.74 | 9.30 |
| 0.50 | 0.45 | 10.67 | 13.60 | 11.25 | 10.67 |

Table 9 shows a partial success. The estimate follows the true level monotonically with a standard deviation of about 0.05 across draws, and it is biased low by 0.04 to 0.08. At the two ends the online rule matches the best fixed discount, 7.07 m with nothing silent and 9.30 m and 10.67 m at 40% and 50%. In between it lags: at 10% and 20% the low bias makes it pick a discount one step weaker than the rule intends in 7 and 11 of the 20 draws, and it ends 0.2 m to 0.3 m behind the fixed $\lambda = 0.25$. The bias has a reading of its own. With no simulated outage, $\hat{q}$ for the later campaign is negative, which says that its phones miss less of the rank-weighted mass of their neighbours than a held-out training phone does; the access points that vanished between the campaigns were weak ones, low in the ranking, and this is consistent with the discount mattering so much for raw levels and so little for ranks in the first column of Table 7. Recalibrating the line on a sample from the target period would remove the bias. We did not do so, since it would have meant touching the later campaign.

### 5.7 H4: self-assessment

Neighbour spread does carry information about the error. Its Spearman correlation with the error of the rank locator is 0.38 under LODO and 0.34 on the later campaign, against 0.25 for the nearest-neighbour distance and none (−0.01) for the number of access points heard. Figure 8(a) shows the relation and its flaw. Mean error grows with spread over most of the range, from 4.9 m to 11.9 m, but the eighth of queries with the smallest spread, 0.3 m on average, has a mean error of 6.9 m and a 90th percentile of 16.5 m, as bad as the queries with seven metres of spread. These are fixes whose neighbours agree on the wrong place.

That flaw defeats the spread-scaled radius (Table 10). Calibrated under LODO for a target of 80%, it covers 84.7% of later-campaign fixes overall, but only 67.8% in the third of queries where it is most confident and 95.7% where it is least; its mean radius is 14.0 m and its pinball loss 2.63, both worse than the constant radius of 12.7 m (loss 2.25). At a 90% target the same happens, 80.3% coverage in the confident third. H4 as stated is refuted. Scaling by spread makes the radius worse than ignoring the query altogether.

**Table 10.** Error radii calibrated with phones held out and scored on the later campaign

| Target | Radius | Coverage (%) | Radius (m) | Pinball |
|---|---|---|---|---|
| 80% | Constant | 85.2 | 12.7 | 2.25 |
| | Spread-scaled | 84.7 | 14.0 | 2.63 |
| | Four-cue quantile | 86.0 | 12.3 | 1.95 |
| 90% | Constant | 94.0 | 18.6 | 1.71 |
| | Spread-scaled | 92.6 | 25.6 | 2.53 |
| | Four-cue quantile | 94.5 | 17.2 | 1.42 |

The four-cue quantile model is the remedy. It lowers the pinball loss by 13% and 17% against the constant radius at the two targets, with a smaller mean radius, and its coverage across the spread terciles is the most even of the three (86.2%, 89.7% and 82.2% at the 80% target, where the constant radius gives 87.8%, 90.5% and 77.4%). All three radii over-cover on the later campaign by three to six points. The reason is visible in Table 2: calibration errors come from LODO, which is harder than the later campaign, so the shift happens to run in the safe direction here. Nothing guarantees that it always will.

### 5.8 Comparison with published results

**Table 11.** UJIIndoorLoc validation results: prior studies and this work

| Method | Floor (%) | Error (m) |
|---|---|---|
| Scalable DNN [14] | 91.27 | 9.29 |
| CNNLoc [15] | 96.03 | 11.78 |
| Hierarchical RNN [16] | 95.23 | 8.62 |
| Local SCNN [17] | 95.52 | 9.68 |
| MG-HGNN [18] | 94.24 | 7.83 |
| kNN, raw level, $\lambda = 1$ | 89.6 | 8.85 |
| kNN, raw level, $\lambda = 0.25$ | 93.6 | 7.21 |
| kNN, rank decay, $\lambda = 1$ | 95.4 | 7.07 |
| kNN, rank decay, $\lambda = 0.25$ | 95.5 | 7.11 |

Table 11 sets the later-campaign figures beside those reported for learned models on the same file. The comparison is loose. Some papers report a three-dimensional error, some tune on the validation file itself, and none was run by us. With those reservations, a nearest-neighbour locator on rank fingerprints,

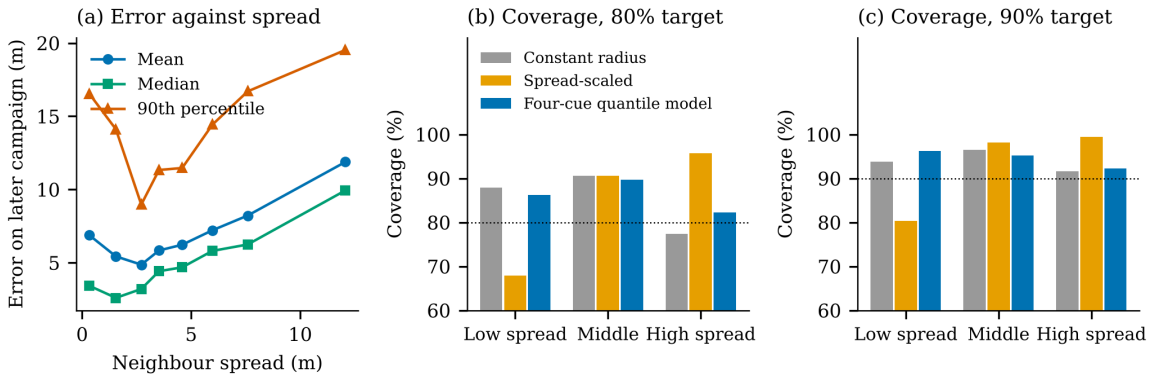



**Figure 8.** Later-campaign error of the rank locator against neighbour spread, in eight equal-count bins (a); coverage of three error radii by tercile of spread at a target of 80% (b) and 90% (c).

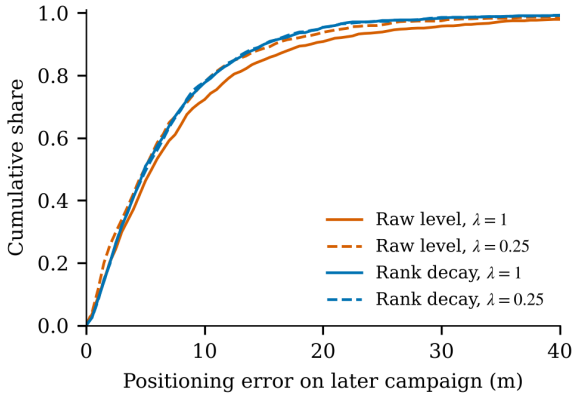



**Figure 9.** Cumulative error distribution on the later campaign for raw-level and rank-decay fingerprints with and without the absence discount.

with one parameter chosen without sight of the test campaign, is at 7.07 m and 95.4%, inside the range of the neural models and ahead of most of them on error. We read this less as a point in favour of nearest neighbours than as a sign that representation and absence handling account for much of what separates the entries of the table.

Figure 9 shows the full error distributions of the four nearest-neighbour configurations. The curves separate above the median, where the wrong-floor and wrong-wing fixes live.

## 6. GUIDANCE FOR DEPLOYMENT

Five pieces of advice follow from the results, in the order in which they cost effort. Report accuracy with locations grouped at the least, and with a device held out wherever more than one device took part in the survey; a random split of this dataset describes a system that is asked only about surveyed points by the phones that surveyed them. Add the absence discount. It needs no retraining and doubles a query time that is well under a millisecond, and at $\lambda = 0.25$ it lowered the later-campaign error of an otherwise unmodified raw-level map from 8.85 m to 7.21 m. Prefer ranks, or the hybrid once it has been confirmed, when the population of phones is unknown and floors matter, which is the usual case in a public building. Keep raw levels when the fleet is fixed, as with tags that all carry one radio. Do not expect offset calibration to help. Finally, size the error radius from several cues and calibrate it with devices held out: under random folds three quarters of the rank locator's errors are below 5.1 m, on the later campaign below 9.3 m, and a radius calibrated on the former would be too small by about that ratio.

## 7. THREATS TO VALIDITY

*Confounding of phone and place.* The dataset was not designed for a device study. Each phone covered some buildings and floors and not others, building 0 has two phones, and nearly every phone was carried by a single user, so that LODO holds out a phone, a user and a set of places at once. The shared-location analysis of Section 5.2 separates place from phone for horizontal error. It cannot separate phone from user, and how a person holds a handset may well matter.

*One benchmark.* All four verdicts rest on one campus in 2013 and on the equipment of that period. Newer deployments have denser and more dynamic access-point sets, band steering and randomised addresses, and the balance between differences in reported values and differences in detection sets may not be the one found here on this campus.

*Simulated outage.* Access points were silenced independently and at random, or by strength. Real failures are correlated, a switch or a floor at a time, and the targeted experiment suggests that clustered loss near the user is much worse than the random case simulated in most of our runs. The natural loss between the two campaigns is real but small.

*Selection effects.* The number of neighbours was chosen on the LODO folds that Table 2 reports, over five candidates. The effect on LODO figures is small next to the differences discussed, and the later campaign was not used for any choice, including $\tau$, the span of 60 dB in Equation (4) and the settings of the quantile model. The hybrid of Section 5.4 is the exception and is labelled as such.

*Statistical power.* Sixteen phones are few. A per-phone difference in horizontal error as large as the pooled one could exist and go undetected at this size, and the claim made for H2 is limited accordingly.

## 8. CONCLUSION

Table 12 collects the verdicts. The optimism of a random split on this benchmark is mostly a matter of replicated locations, and the specific cost of a new phone is the floor. Rank fin-

gerprints repair a good part of that cost and offset correction none of it, in keeping with responses whose slopes differ. With $m$ the number of access points a weak phone still hears, the rank fingerprint moves by less than

$$\left.\frac{e^{-2m/\tau}}{1-e^{-2/\tau}}\right|_{\tau=5,\ m=10} \approx 0.056 \tag{16}$$

in squared norm, under 2% of its own, which is why the gain concentrates on queries with few access points. A single discount on map-only access points keeps the error within 21% of its clean value when three access points in ten are silent, and on raw levels it closes most of the gap to ranks before any outage is simulated. Neighbour spread should not be used alone to size an error radius. An exploratory hybrid of levels and ranks improved 14 of the 16 phones, and it deserves a planned test.

**Table 12.** Verdicts on the four hypotheses

| | |
|---|---|
| **H1** | Supported. Phone held out: 10.07 m against 2.21 m; 1.72 m and 10.6 floor points are specific to the device. |
| **H2** | Supported for floors (12 of 16 phones, $p = 0.021$) and for the later campaign ($-1.78$ m); not shown per phone for horizontal error ($p = 0.30$). Offset correction: no gain. |
| **H3** | Supported. At 30% outage, 8.62 m against 19.18 m; premium of 0.7 m under LODO with no outage. |
| **H4** | Refuted as stated. Spread-scaled radius covers 67.8% of its most confident third at an 80% target; a four-cue quantile model lowers pinball loss by 13% to 17%. |

What remains open is practical. The best discount depends on an outage level that a deployment does not know, and the online estimate tried here follows that level with a bias it cannot correct by itself; a calibration sample from the period of use would settle whether the bias is all that separates the online rule from the best fixed choice. The device study should also be repeated on a collection built for it, with every phone at every point.

## REFERENCES


[1] P. Bahl and V. N. Padmanabhan, “RADAR: an in-building RF-based user location and tracking system,” in Proc. IEEE INFOCOM, pp. 775–784, 2000.

[2] M. Youssef and A. Agrawala, “The Horus WLAN location determination system,” in Proc. 3rd International Conference on Mobile Systems, Applications, and Services (MobiSys), pp. 205–218, 2005.

[3] A. Haeberlen, E. Flannery, A. M. Ladd, A. Rudys, D. S. Wallach, and L. E. Kavraki, “Practical robust localization over large-scale 802.11 wireless networks,” in Proc. ACM MobiCom, pp. 70–84, 2004.

[4] M. B. Kjærgaard, “Indoor location fingerprinting with heterogeneous clients,” Pervasive and Mobile Computing, vol. 7, no. 1, pp. 31–43, 2011.

[5] C. Laoudias, M. P. Michaelides, and C. G. Panayiotou, “Fault detection and mitigation in WLAN RSS fingerprint-based positioning,” Journal of Location Based Services, vol. 6, no. 2, pp. 101–116, 2012.

[6] G. M. Mendoza-Silva, P. Richter, J. Torres-Sospedra, E. S. Lohan, and J. Huerta, “Long-term WiFi fingerprinting dataset for research on robust indoor positioning,” Data, vol. 3, no. 1, Art. no. 3, 2018.

[7] J. Torres-Sospedra et al., “UJIIndoorLoc: a new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems,” in Proc. International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 261–270, 2014.

[8] S. He and S.-H. G. Chan, “Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons,” IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 466–490, 2016.

[9] F. Zafari, A. Gkelias, and K. K. Leung, “A survey of indoor localization systems and technologies,” IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2568–2599, 2019.

[10] T. Cover and P. Hart, “Nearest neighbor pattern classification,” IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, 1967.

[11] J. Torres-Sospedra, R. Montoliu, S. Trilles, O. Belmonte, and J. Huerta, “Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems,” Expert Systems with Applications, vol. 42, no. 23, pp. 9263–9278, 2015.

[12] J. Torres-Sospedra et al., “A realistic evaluation of indoor positioning systems based on Wi-Fi fingerprinting: the 2015 EvAAL-ETRI competition,” Journal of Ambient Intelligence and Smart Environments, vol. 9, no. 2, pp. 263–279, 2017.

[13] M. Nowicki and J. Wietrzykowski, “Low-effort place recognition with WiFi fingerprints using deep learning,” in Automation 2017, Advances in Intelligent Systems and Computing, vol. 550. Cham, Switzerland: Springer, pp. 575–584, 2017.

[14] K. S. Kim, S. Lee, and K. Huang, “A scalable deep neural network architecture for multi-building and multi-floor indoor localization based on Wi-Fi fingerprinting,” Big Data Analytics, vol. 3, Art. no. 4, 2018.

[15] X. Song et al., “A novel convolutional neural network based indoor localization framework with WiFi fingerprinting,” IEEE Access, vol. 7, pp. 110698–110709, 2019.

[16] A. E. A. Elesawi and K. S. Kim, “Hierarchical multi-building and multi-floor indoor localization based on recurrent neural networks,” arXiv preprint arXiv:2112.12478, 2021.

[17] N. Agah, B. Evans, X. Meng, and H. Xu, “A local machine learning approach for fingerprint-based indoor localization,” arXiv preprint arXiv:2302.10810, 2023.

[18] Y. Wang, Z. Zhang, N. Li, X. Zhao, D. Zhao, and T. Zhao, “MG-HGNN: a heterogeneous GNN framework for indoor Wi-Fi fingerprint-based localization,” arXiv preprint arXiv:2511.07282, 2025.

[19] A. K. M. Mahtab Hossain, Y. Jin, W.-S. Soh, and H. N. Van, “SSD: a robust RF location fingerprint addressing

mobile devices' heterogeneity," IEEE Transactions on Mobile Computing, vol. 12, no. 1, pp. 65–77, 2013.

[20] J. Machaj, P. Brida, and R. Piché, "Rank based fingerprinting algorithm for indoor positioning," in Proc. International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–6, 2011.

[21] S. Zhang, J. Guo, N. Luo, D. Zhang, W. Wang, and L. Wang, "A calibration-free method based on grey relational analysis for heterogeneous smartphones in fingerprint-based indoor positioning," Sensors, vol. 19, no. 18, Art. no. 3885, 2019.

[22] E. S. Lohan, J. Torres-Sospedra, H. Leppäkoski, P. Richter, Z. Peng, and J. Huerta, "Wi-Fi crowdsourced fingerprinting dataset for indoor positioning," Data, vol. 2, no. 4, Art. no. 32, 2017.

[23] H. Lemelson, M. B. Kjærgaard, R. Hansen, and T. King, "Error estimation for indoor 802.11 location fingerprinting," in Location and Context Awareness (LoCA), Lecture Notes in Computer Science, vol. 5561. Berlin, Germany: Springer, pp. 138–155, 2009.

[24] V. Vovk, A. Gammerman, and G. Shafer, Algorithmic Learning in a Random World. New York, NY, USA: Springer, 2005.

[25] J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman, "Distribution-free predictive inference for regression," Journal of the American Statistical Association, vol. 113, no. 523, pp. 1094–1111, 2018.

[26] A. N. Angelopoulos and S. Bates, "Conformal prediction: a gentle introduction," Foundations and Trends in Machine Learning, vol. 16, no. 4, pp. 494–591, 2023.

[27] K. A. Nguyen, "A performance guaranteed indoor positioning system using conformal prediction and the WiFi signal strength," Journal of Information and Telecommunication, vol. 1, no. 1, pp. 41–65, 2017.

[28] M. C. Molina, M. V. Da Silva, I. Ahriz, R. D. Souza, and H. Alves, "Indoor Localization for Navigation Applications using Conformal Prediction," in Proc. 16th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), Yorktown Heights, NY, USA, pp. 0574–0580, 2025.

[29] R. J. Tibshirani, R. F. Barber, E. J. Candès, and A. Ramdas, "Conformal prediction under covariate shift," in Advances in Neural Information Processing Systems, vol. 32, 2019.

[30] T. S. Rappaport, Wireless Communications: Principles and Practice, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2002.

[31] R. Koenker and G. Bassett, "Regression quantiles," Econometrica, vol. 46, no. 1, pp. 33–50, 1978.

[32] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," The Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.

[33] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Machine Learning, vol. 63, no. 1, pp. 3–42, 2006.

[34] F. Pedregosa et al., "Scikit-learn: machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[35] F. Wilcoxon, "Individual comparisons by ranking methods," Biometrics Bulletin, vol. 1, no. 6, pp. 80–83, 1945.

[36] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," Journal of Machine Learning Research, vol. 7, pp. 1–30, 2006.

[37] B. Efron and R. J. Tibshirani, An Introduction to the Bootstrap. New York, NY, USA: Chapman & Hall, 1993.

[38] D. R. Roberts et al., "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure," Ecography, vol. 40, no. 8, pp. 913–929, 2017.