City-Wide Signal Strength Maps: Prediction with Random Forests

E. Alimpertis*, A. Markopoulou*, C. T. Butts*, K. Psounis+
University of California, Irvine* Univ. of Southern California+

LTE, UCI-Mobile

Mobile is King

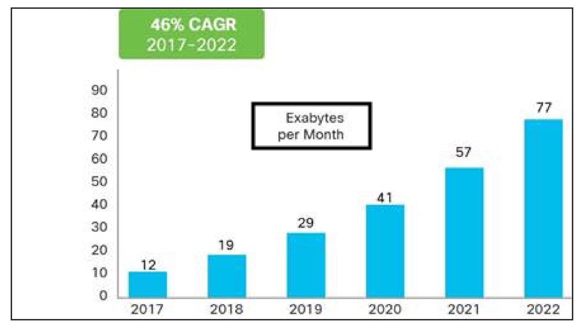


Figure 1: Exponential Global Mobile Data Traffic Growth, Source[1].

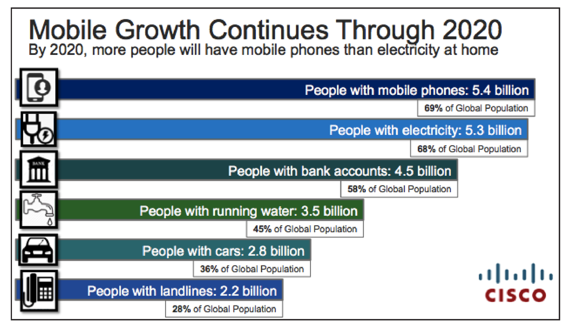


Figure 2: More People with Mobile than Running Water, Source[1].

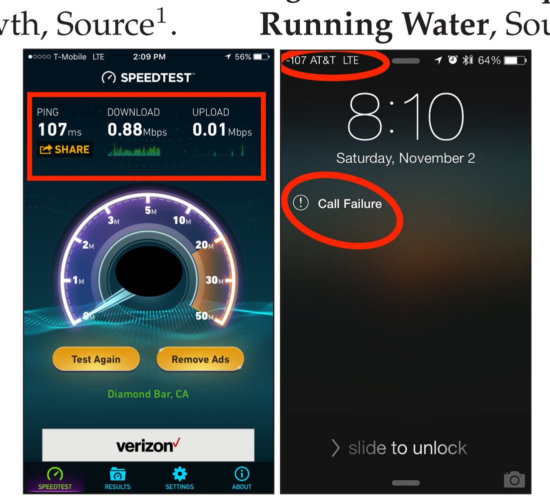


Figure 3: However, all of us have experienced: **Poor Performance** and **Failed Calls**.

1. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017-2022

Signal Strength Maps Overview

- **Mobile Signal (Coverage) Maps** by **Users' Mobiles** as Sensors!
  - ✓**Large Scale**, ✓**Long Periods**.
  - **However:** Measurements are ✗sparse, ✗inadequate, ✗expensive.
- Cellular Providers Signal Maps Data:
  1. Themselves (wardriving, privacy concerns).
  2. Mobile Analytics Companies ($$), *e.g.*, OpenSignal, Tutela, RootMetrics.

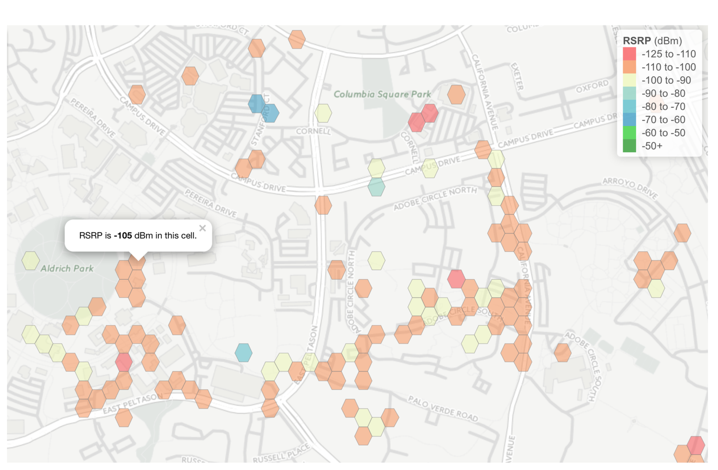


Figure 4: T-Mobile LTE Map of UCI, Collected by Ourselves [1].

Goals and Contributions

Goal: How to **Predict** missing values in space, time and other features?
Benefits and Contributions: Cheaper & More Accurate Maps for:

1. **Users:** Find Best Network.
2. **Cellular** Carriers:
   (a) Monitor their and competitors' Cellular Network.
   (b) Network Management and Upgrades (*e.g.*, Deploy more cells).
   (c) SDN/SON *e.g.*, network selection *etc.*
3. **Mobile analytics Companies**: Reduce Operational Costs (*e.g.*, AWS cost).

Signal Strength Maps Prior Work

	Features			Setup: Environment, Scale and Data		
	Spatial	Time	Device Network	Environment Agnostic	City-Wide	No Expensive LiDar Data
Log-Distance Path-Loss ($LDPL$) [2]	✓	✗	✗	✓	✗	✓
COST-231/ WINNER I-II/ Ray Tracing	✓	✗	✓	✗	✗	✓
Geostatistics SpecSense [3]	✓	✗	✗	✗	✗	✓
BCS [4]	✓	✓	✗	✗	✗	✓
RAIK-DNNs [5]	✓	✗	✗	✗	✓	✗
Our Work: Random Forests	✓	✓	✓	✓	✓	✓

Table 1: Signal Maps Approaches **Compared with Our Work.**

Random Forests (RFs) LTE RSRP Prediction

- LDPL RSRP modeling: $P_{cID}^{(t)}\left(\vec{l_j}\right) = P_0^{(t)} - 10n_j \log_{10}\left(||\vec{l}_{BS} - \vec{l_j}||_2/d_0\right) + w_j^{(t)}$.
- **RSRP Prediction:** $\widehat{P} \sim \mathcal{N}(RFs_\mu(\mathbf{x}), \sigma_\mathbf{x}^2)$: P_j target, $\mathbf{x}_j$ feature vector.
- Random Forests (RFs) are an ensemble of multiple decision trees.
- **RSRP Predictors:**
  1. RFs$_{x,y}$: Spatial Only (localization in [6]).
  2. RFs$_{x,y,t}$: Spatiotemporal Features.
  3. RFs$_{all}$: All Features.
- For each measurement P_j we consider the full set of features:
- **Features** $\mathbf{x}_j$: $\mathbf{x}_j^{\mathbf{full}} = (l_j^x, l_j^y, d, h, cID, dev, out, ||\vec{l}_{BS} - \vec{l_j}||_2, freq_{dl})$
  1. **Location**, (l_j^x, l_j^y).
  2. **Time Features**, $\mathbf{t_j} = (d, h)$. RSRP Variance Is Time Dependent.
  3. **Cell-ID**, cID. RSRP is defined per serving cell.
  4. **Device Hardware Type**, dev. RSRP calculation differs per device/hardware differences.
  5. **Downlink Carrier Frequency**, $freq_{dl}$. Radio propagation depends on $freq_{dl}$.
  6. **Outdoors**, out. From Android's API GPS velocity.
  7. **Distance** between UE – BS, $||\vec{l}_{BS} - \vec{l_j}||_2$.

Why RFs for Data-Driven Prediction?

1. RFs inherently **considers all features** x; Geostatistics [3] **only spatial**.
2. RFs Automatically **identifies areas** with **spatially (and temporal) correlated RSRP** (similar wireless propagation characteristics).

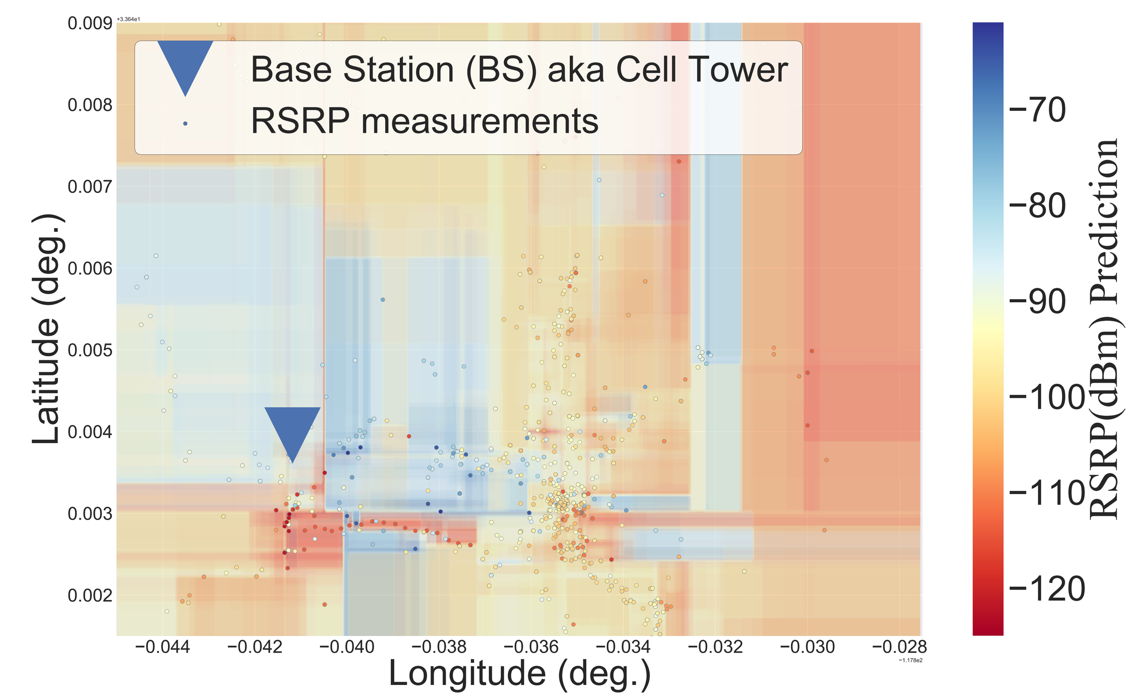


Figure 5: Example of decision boundaries chosen by RFs$_{x,y}$ for Campus cell x306.

Datasets

Dataset	Period	Areas	Type of Measurements	Characteristics	Source
Campus	02/10/17 - 06/18/17	Univ. Campus Area $\approxeq 3km^2$	LTE KPIs: RSRP, [RSRQ]. Context: GPS Location, timestamp, dev, cid. Features: $\mathbf{x} = \left(l_j^x, l_j^y, d, h, dev, out, ||\vec{l}_{\mathrm{BS}} - \vec{l_j}||_2\right)$	No. Cells = 25 No. Meas $\approxeq$ 180K Density ($\frac{N}{m^2}$) Per Cell: 0.01 - 0.66 (Table 3) Overall Density: 0.06	Ourselves [1]
NYC & LA	09/01/17- 11/30/17	NYC Metropolitan Area $\approxeq 300km^2$ LA metropolitan Area $\approxeq 1600km^2$	LTE KPIs: RSRP, [RSRQ, CQI]. Context: GPS Location, timestamp, dev, cid. EARFCN. Features:$\mathbf{x} = \left(l_j^x, l_j^y, d, h, cid, dev, out, ||\vec{l}_{\mathrm{BS}} - \vec{l_j}||_2, freq_{dl}\right)$	No. Meas NYC $\approxeq$ 4.2M No. Cells NYC $\approxeq 88k$ Density NYC-all $\approxeq 0.014\frac{N}{m^2}$ No. Meas LA $\approxeq$ 6.7M No. Cells LA $\approxeq$ 111K Density LA-all $\approxeq 0.0042\frac{N}{m^2}$	Mobile Analytics Company

Table 2: Overview of Signal Maps Datasets used in this study

	NYC Manhattan Midtown (Fig.8)	NYC E. Brooklyn	LA Southern
MN-Carrier	MNC-1	MNC-1	MNC-2
No. Measurements	$\approxeq$ 63K	$\approxeq$ 104K	$\approxeq$ 20K
Area km^2	$1.8km^2$	$44.8\ km^2$	$220\ km^2$
Data Density $\frac{N}{m^2}$	$\approxeq$ 0.035	$\approxeq$ 0.002	$\approxeq$ 0.0001
No. Cells $\lvert C\rvert$	429	721	353
Cell Density $\frac{\lvert C\rvert}{km^2}$	238.3	16.1	1.6

Table 3: NYC and LA datasets: LTE TAs Examples.

Datasets Examples and Feature Importance

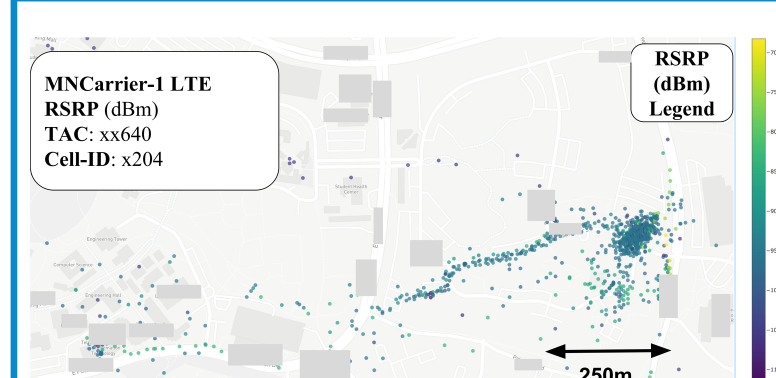


Figure 6: Campus example cell x204: high density (0.66), low dispersion (325).

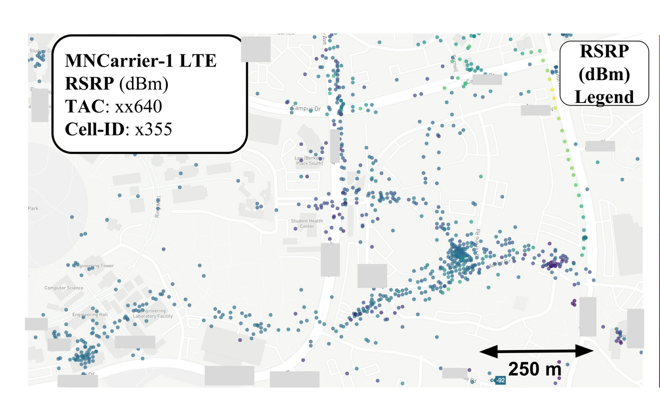


Figure 7: Campus: example cell x355: small density (0.12) more dispersed data (573).

Figure 8: NYC: Manhatta Midtownn LTE TA

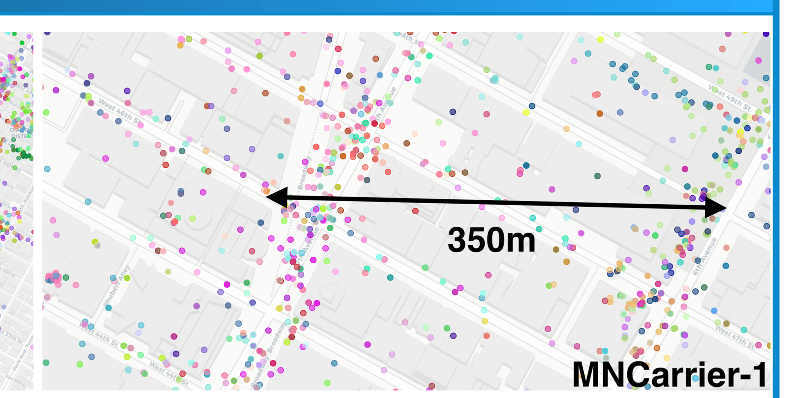
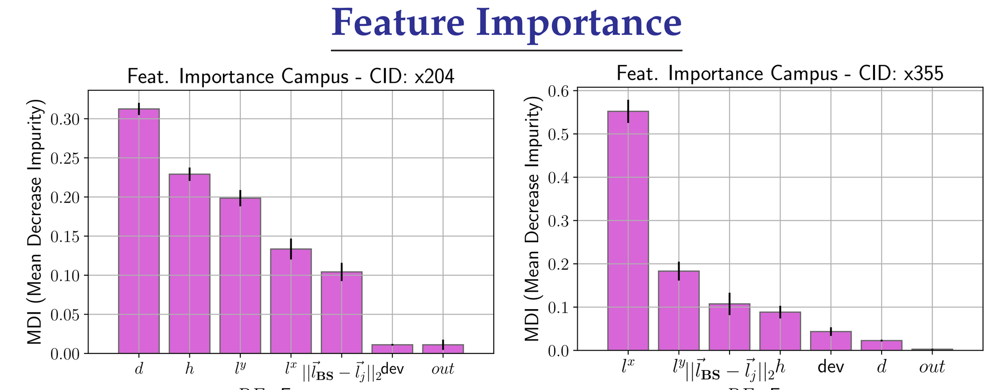


Figure 9: NYC: zooming in Manhattan Midtown (Time Square) for some of the available cells (Different color per cID).

Feature Importance



(1) Model – Based (Wireless Propagation)	(a) $LDPL$	**(b) LDPL-knn** (Heteregoneous n_j)
(2) Geospatial Interpolation	(a) OK Ordinary Kriging	(b) OKD OK Detrening
(3) Random Forests (RFs**)**	(a) $RFs_{x,y}$	**(b)** $RFs_{x,y,t}$ **(c)** RFs_{all}

Table 4: Evaluation: Methods For Comparison

RFs Setup	Model Per LTE TA	Model per cID
RFs hyperparameters	$n_{trees} = \{20, 100\}$	$\max_{depth} = \{20, 30\}$

Table 5: RFs **Setup and Hyperparameters.**

Comparing Predictors per cell

	Cell Characteristics					RMSE (dB)						
cID	N	$\frac{N}{\text{sq}\ m^2}$	SDD	$\mathbb{E}[P]$	σ^2	LDPL hom	LDPL kNN	OK	OKD	RFs x,y	RFs x,y,t	RFs all
x914	3215	0.007	791	-94.5	96.3	13.3	3.47	3.59	2.28	3.43	1.71	**1.67**
x034	1564	0.010	441	-101.2	337.5	19.5	7.82	7.44	5.12	7.56	3.82	**3.84**
x901	16051	0.162	355	-107.9	82.3	8.9	4.60	4.72	3.04	4.54	1.73	**1.66**
x204	55566	0.666	325	-96.0	23.9	6.9	3.84	3.85	2.99	3.83	2.30	**2.27**
x922	3996	0.107	218	-102.7	29.5	5.6	3.1	3.16	2.01	3.10	1.92	**1.82**
x902	34193	0.187	481	-111.5	8.1	21.0	2.60	2.47	1.64	2.50	1.37	**1.37**

	Cell Characteristics					RMSE (dB)						
cID	N	$\frac{N}{\text{sq}\ m^2}$	SDD	$\mathbb{E}[P]$	σ^2	LDPL hom	LDPL kNN	OK	OKD	RFs x,y	RFs x,y,t	RFs all
x470	7699	0.034	533	-107.3	16.9	24.8	3.64	2.73	1.87	2.78	1.26	**1.26**
x915	4733	0.042	376	-110.6	203.9	14.3	7.54	7.39	4.25	7.31	3.29	**3.15**
x808	12153	0.035	666	-105.1	7.7	4.40	2.41	2.42	1.60	2.34	1.75	**1.59**
x460	4077	0.040	361	-88.0	32.8	11.2	2.35	2.28	1.56	2.31	1.84	**1.84**
x306	4076	0.011	701	-99.2	133.3	18.3	4.85	4.30	2.80	3.94	3.1	**3.06**
x355	30084	0.116	573	-94.3	42.6	9.3	2.42	2.31	1.85	2.26	1.79	**1.79**

UCI Networking Group Webpage

Networking Group@UCI: `athinagroup.eng.uci.edu`

Model Granularity: cID vs. LTE TA

Question: At What *Level of Granularity* Should we Train Our RFs Models?

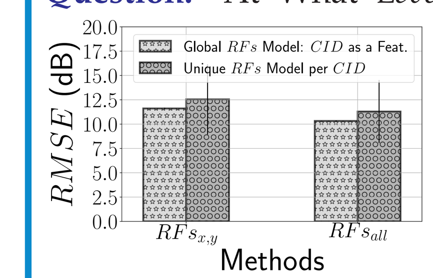


Figure 10: MNC-1, Manhattan Midtown (urban).

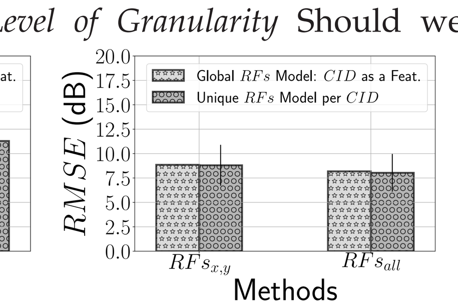


Figure 11: MNC-1, East Brooklyn (suburban)

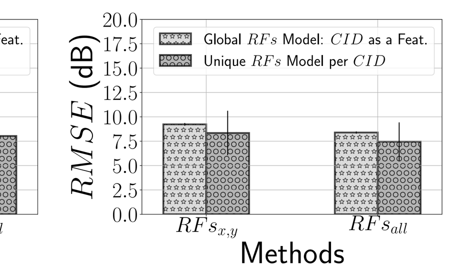


Figure 12: MNC-2, Southern LA (suburban).

Number of Measurements vs. $RMSE$ Trade-off

Tradeoffs: (1) **80% Less Data:** Same Accuracy.
(2) Same Data **17% Relative Error Reduction** (or 1dB error reduction).

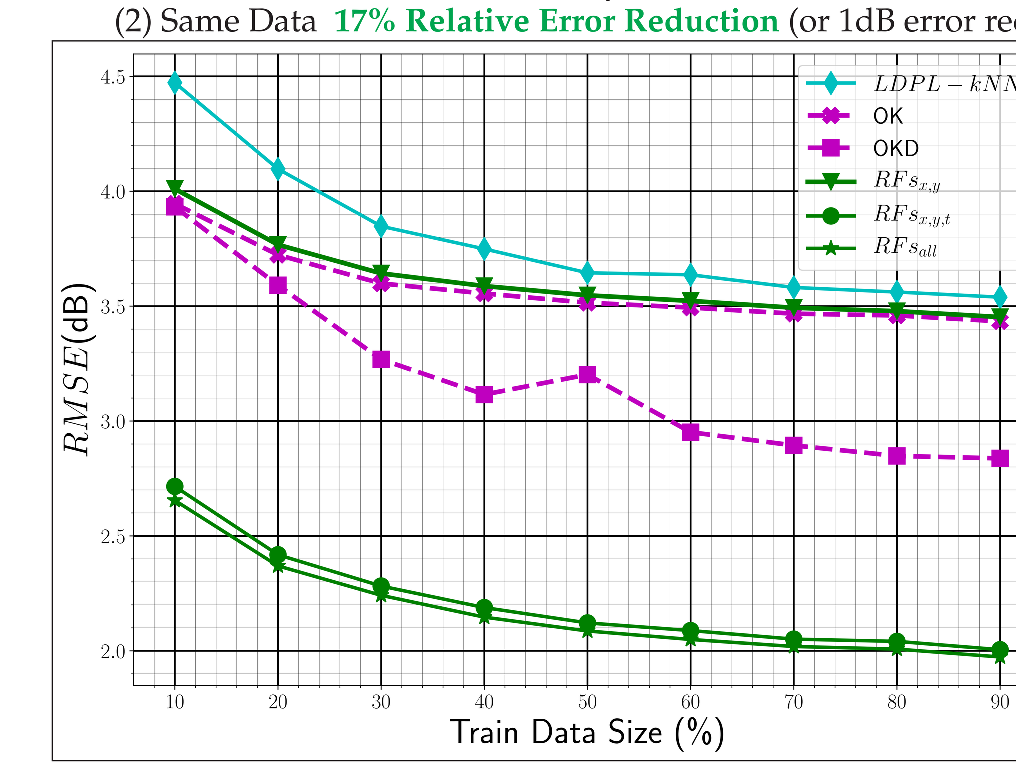


Figure 13: Campus dataset: $RMSE$ vs. Training Size. Our methodology (RFs with more than spatial features, *i.e.*, RFs$_{x,y,t}$, RFs$_{all}$) significantly improves the RMSE-cost tradeoff: it can reduce $RMSE$ by 17% for the same number of measurements compared to state-of-the-art data-driven predictors(OKD); *or* it can achieve the lowest error possible by OKD ($\simeq$ 2.8dB) with 10% instead of 90% (and 80% reduction) of the measurements.

City-Wide Maps: NYC and LA datasets Experiments

CDFs for $RMSE$ per cID for two different LTE TA, for the same major MNC-1.
Benefits: (1) RFs$_{all}$ offer **2dB gain** over the baselines for the 90th percentile.
(2) 2dB for 1-bar in voLTE means 1-5% call drop rate [7]

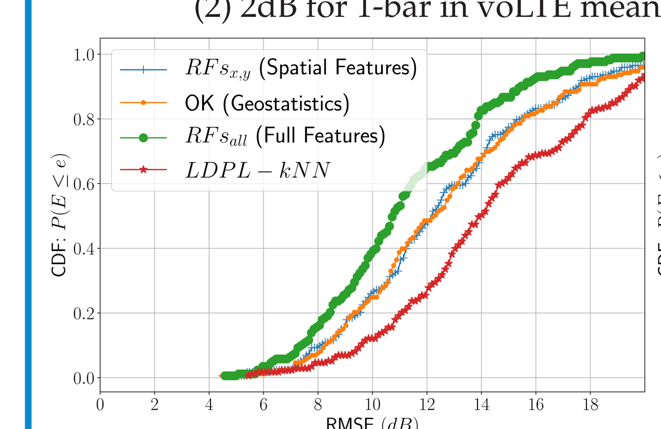


Figure 14: MNC-1, NYC Manhattan Midtown (urban).

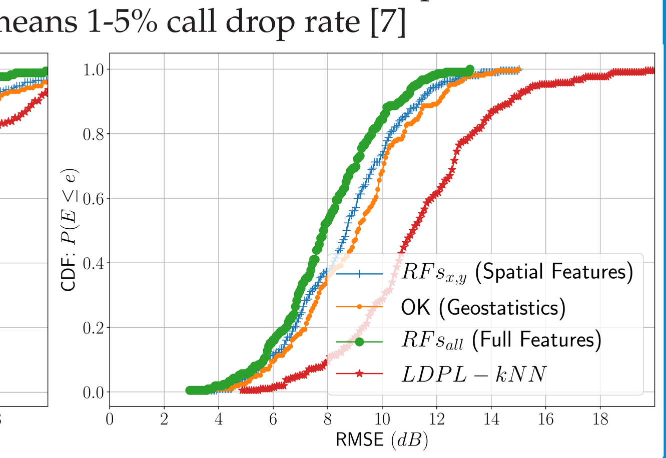


Figure 15: MNC-1, LA Southern (Suburb)

References

[1] E. Alimpertis and A. Markopoulou. A system for crowdsourcing passive mobile network measurements. In *14th USENIX NSDI'17, Posters Sessions*, Boston, Massachusetts, USA, March 2017. USENIX Association.

[2] E. Alimpertis, N. Fasarakis-Hilliard, and A. Bletsas. Community RF Sensing for Source Localization. *IEEE Wireless Commun. Lett.*, 3(4):393–396, Aug 2014.

[3] A. Chakraborty et al. Specsense: Crowdsensing for efficient querying of spectrum occupancy. In *Proc. of the IEEE INFOCOM*, Atlanta, Georgia, USA, May 2017.

[4] S. He and K. G. Shin. Steering crowdsourced signal map construction via bayesian compressive sensing. In *Proc. of the IEEE INFOCOM '18*, pages 1016–1024. IEEE, April 2018.

[5] R. Enami et al. RAIK: Regional analysis with geodata and crowdsourcing to infer key performance indicators. In *Proc. of the IEEE WCNC '18*, pages 1–6, April 2018.

[6] A. Ray et al. Localization of lte measurement records with missing information. In *Proc. of the IEEE INFOCOM*, pages 1–9, San Fransisco, CA, USA, April 2016.

[7] Yunhan Jack et al. Jia. Performance characterization and call reliability diagnosis support for voice over LTE. In *Proc. of the ACM MobiCom*, pages 452–463, 2015.

Acknowledgments

This work was supported by NSF Awards 1649372, 1526736, 1444060; as well as by Networked Systems and CPCC at UC Irvine and the Broadcom Foundation Fellowship program.