Georeferenced Spectrum Occupancy Analysis Using Spatially Very Sparse Monitoring Data

Serhat Tadik[1], Neal Patwari[2], Kirk Webb[3], Xuan Lin[1], Gregory D. Durgin[1]

[1]Georgia Institute of Technology   [2]Washington University in St. Louis   [3]University of Utah

Abstract—The complex and invisible nature of dynamic radio usage poses a significant challenge, even for the most skilled engineers, in visualizing and effectively utilizing radio frequency data. This paper proposes a method to quantify and visualize what is known about spectrum occupancy in a region from spectrum monitoring data recorded at a few fixed locations. The average observed signal strengths at the monitors are extrapolated throughout the region through likelihood estimation of transmitter location(s) and a simple log-distance path loss model. New georeferenced spectrum occupancy visualizations that combine estimates of occupancy power with duty cycle, and of signal variation with confidence level are introduced offering insights into planning for future allocations, interference, and broadcast coverage analysis. The spectrum consumption trends over different times of day and seasons are analyzed and interpreted.

Index Terms—digital spectrum twins, spectrum occupancy

I. Introduction

With the increasing demand for wireless services, the radio spectrum—a finite resource—is experiencing pressure to support a diverse array of services, from cellular networks to satellite communications. The inherent challenge lies in the transient and invisible nature of radio frequency (RF) usage, further complicated by the complex dynamics of the RF environment. These elements present considerable difficulties in effectively visualizing and optimizing this invaluable resource, which is crucial for maintaining and advancing wireless communication systems.

Understanding and effectively managing the RF spectrum requires not only the ability to sense and monitor spectrum usage but also the capacity to visualize and interpret this data in a meaningful way. This is where the concept of a *digital spectrum twin* (DST) becomes important. A DST is a dynamic, cloud-based model of the RF spectrum. It continuously tracks and analyzes radio activity within a specific geographic area through propagation measurements and modeling, supported by advanced parallel intelligence algorithms [1]–[3]. At the heart of a DST are georeferenced maps, consisting of propagation picture elements, or *proxels*. These proxels digitally represent the geographic region, offering a clear and accessible way to visualize and understand RF activity in the area. This new architecture simplifies the process of interpreting complex RF data, making it easier to manage and optimize spectrum use for identifying underutilized frequencies in certain parts of
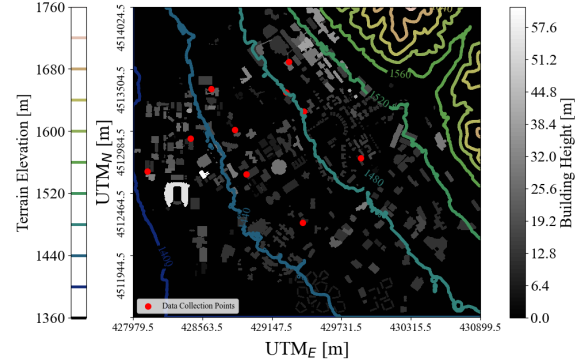
Fig. 1: Terrain profile, building layout, and ground-level fixed monitoring locations, University of Utah campus

a region, detecting unauthorized transmissions, and planning for future spectrum allocations.

This paper addresses the spectrum occupancy visualization and interpretation challenge by adopting the DST architecture and visualizing the estimated spectrum occupancy in a specific geographical region, using a limited number of fixed-location sensors. In our paper, we show results using data from 10 fixed spectrum monitors in the 2.6 km by 2.9 km area shown in Figure 1 of the POWDER testbed [4] over an extensive period of eleven months at 2.1 GHz, C, and CBRS bands.

The contributions of this paper can be summarized as:

- a likelihood estimation technique for extrapolating spatially very sparse average occupancy power observations throughout a region,
- a technique for extrapolating spatially very sparse signal variation observations leveraging the nonlinear relationship between occupancy power and signal variation,
- novel georeferenced visualizations illustrating extrapolated power & duty cycle and signal variation & confidence level in combined plots,
- daily and seasonal analysis of spectrum usage in a region and interpretation of the trends.

II. Methodology

Our overall goal is to estimate the received power at all locations in an area and all frequency bands of interest, given the m spectrum measurements made by sensors at known locations. We assume the transmitter locations are not known. Further, we don't know how many transmitters there might be. In our method, we consider them to be nuisance parameters –
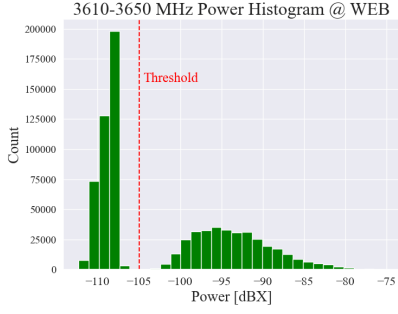
Fig. 2: Power Level Histogram at 3610-3650 MHz

they are needed to estimate received powers but are otherwise unimportant to our goal. In our method, we discretize the area into N proxels and assume that a transmitter (with unknown transmit power) can be in any proxel. This formulation avoids explicitly estimating the number of transmitters, their locations, or their transmit power. Our discretized transmit power image estimates the transmit power emitted from each proxel. As we describe, these intermediate nuisance parameters are estimated first and then used to estimate the received power field, which is ultimately our goal.

A. Experiment

RF compliance data is collected at fixed monitoring stations within the POWDER platform with a median time difference of approximately 6.5 hours between observations. The monitors sweep a wide range of frequencies ranging from 86 to 6011 MHz with a step size of 60 kHz. The collected data is a dictionary of frequency and received signal strength indicator (RSSI) pairs.

1) Occupancy vs. Noise Power: The data collected at 10 fixed monitoring stations over 11 months is analyzed. The power level histograms are generated as illustrated in Figure 2, and a threshold separating occupancy and noise is determined by analyzing both the histogram and the overall frequency domain signal power graph. Thresholding was selected for spectrum usage detection as the absence of complex receiver data precluded more advanced methods such as waveform-based or spectral correlation based sensing.

After establishing the threshold, the average occupancy power at a monitor is calculated by taking the mean power level in decibels of the data points that exceed the threshold. Similarly, the duty cycle is computed as the proportion of data points surpassing the threshold relative to the total number of data points in the dataset. Finally, signal variation is calculated by taking the variance of the power level in decibels of the data points that exceed the threshold.

B. Average Occupancy Power

The methodology begins by calculating occupancy power at various sensor locations, with $\boldsymbol{p} \in \mathbb{R}^m$ representing a realization of the random multivariate received power vector $\boldsymbol{P} \in \mathbb{R}^m$ at m spectrum sensors, and p_j denoting the power

received at sensor j. According to the log-distance path loss model [5], the theoretical mean received power at sensor j, denoted as p_j^{th}, is expressed by

$$p_j^{th} = t_i - p_{i0} - 10 n_p \log_{10}(d_{ij}/d_{i0}), \quad (1)$$

where t_i signifies the unknown transmitter power, p_{i0} is a reference power level observed at a distance d_{i0}, n_p is the path loss exponent, and d_{ij} is the distance between the potential transmitter location i and sensor j. All the power values discussed in this paper are indicators of actual power levels with an offset. Therefore, p_{i0} is selected to be 0 at a d_{i0} of 1. Although theoretical and ray-tracing based models could offer a more detailed perspective of propagation in this complex semi-urban environment, the log-distance path loss model was chosen for its simplicity and to establish a foundation for future research.

1) Estimation of Transmit Power: The initial objective is to estimate t_i for each proxel i. This value t_i answers the question, if a transmitter were located in proxel i, what power would it have to be transmitting? Our method answers this question by finding the t_i that minimizes the L_2 norm of the difference between the observed power vector $\boldsymbol{p}$ and a vector of theoretical power levels $\boldsymbol{p}^{th}$, whose elements are predicted by (??) as a function of t_i. In other words, for the error vector,

$$\boldsymbol{e_i}(t_i) = \boldsymbol{p}^{th}(t_i) - \boldsymbol{p}, \quad (2)$$

the estimation of t_i is the optimization problem of minimizing the L_2 norm of the error vector with respect to t_i,

$$\hat{t}_i = \operatorname{argmin} \|\boldsymbol{e_i}(t_i)\|_2. \quad (3)$$

2) Probability Distribution of Transmitter Locations: Let $\hat{\boldsymbol{p}}^{th}(\hat{t}_i) \in \mathbb{R}^m$ denote a vector of optimized theoretical power levels whose elements are predicted by (??) as a function of optimal transmit power $\hat{t}_i$. Accordingly, the error vector between $\boldsymbol{p}$ and $\hat{\boldsymbol{p}}^{th}$ at the potential transmitter location i is denoted as $\hat{\boldsymbol{e}}_{\boldsymbol{i}}(\hat{t}_i)$.

A likelihood function is then constructed to represent the probability of each transmitter location given the observed data. The likelihood for a given location i is assumed to have a multivariate Gaussian distribution considering that the shadowing loss i.e. medium-scale deviations from the large-scale path loss model has a Gaussian distribution in decibels supported by measurements [5], [6] and analytical derivations [7],

$$\begin{aligned} \mathcal{L}(\hat{\boldsymbol{p}}^{th}, \boldsymbol{V}; \boldsymbol{p}) &= f(i|\boldsymbol{p}) \\ &= \frac{1}{\det(\boldsymbol{V})^{\frac{1}{2}}(\sqrt{2\pi})^m} \exp\left(-\frac{1}{2}\hat{\boldsymbol{e}}_{\boldsymbol{i}}^T \boldsymbol{V}^{-1} \hat{\boldsymbol{e}}_{\boldsymbol{i}}\right), \end{aligned} \quad (4)$$

where $f(i|\boldsymbol{p})$ is the conditional probability mass function of transmitter location given observed power vector $\boldsymbol{p}$, and $\boldsymbol{V} \in \mathbb{R}^{m \times m}$ is the covariance matrix. The elements of the covariance matrix are modeled assuming that the loss field has an exponentially decaying spatial correlation [8], [9],

$$\boldsymbol{V}_{kl} = \sigma^2 \exp\left(-d_{kl}/\delta_c\right), \quad (5)$$

(a) Distribution of estimated tx power (b) 2D PMF of transmitter location (c) Estimated signal strength
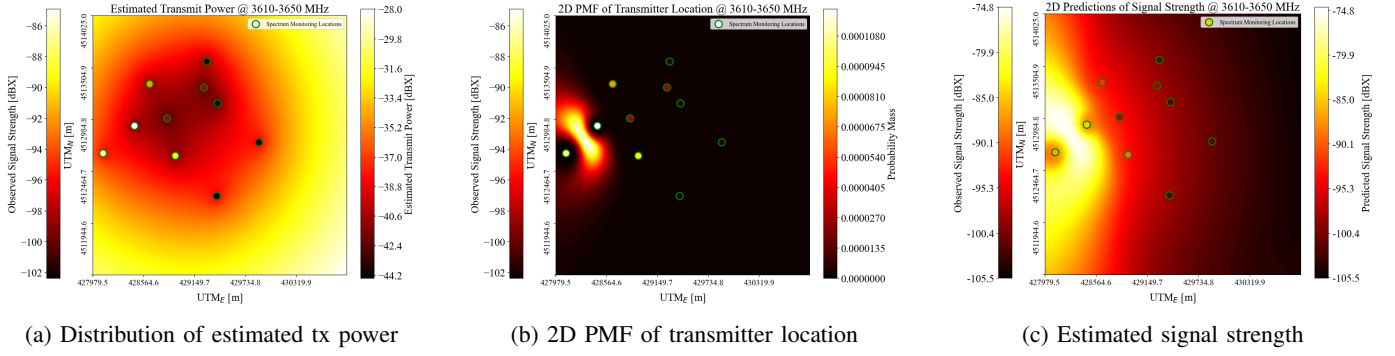
Fig. 3: Signal strength estimation process, 3610-3650 MHz, University of Utah campus

where σ^2 is the variance of the transmit power field, d_{kl} is the distance between observed sensors k and l, and δ_c is the spatial correlation coefficient.

3) Estimation of Received Power at Target Locations:
The final step involves estimating the received power at various locations within the region. For a target location y, the received power is estimated as an expected value, that is, a weighted sum of the received power predicted from each potential transmitter location with its estimated transmit power, weighted by the probability that a transmitter is located there:

$$p_{\text{est}}(j) = \sum_{i=1}^{N} \hat{p}_j^{th}(\hat{t}_i) f(i|\boldsymbol{p}), \qquad (6)$$

where $\hat{p}_j^{th}(\hat{t}_i)$ is the optimized theoretical path loss between the transmitter location i and the target location j as outlined by Equation (**??**), and N is the total number of proxels. If the map area is rectangular, N can be formulated as $\frac{wh}{r_p^2}$ where w is the width and h is the height of the map in meters, and r_p is the width/height of a single proxel in meters.

C. Duty Cycle

The duty cycle of transmissions varies based on the choices of the transmitting authorities, making it unsuitable for parametric distribution modeling. However, there's a noticeable spatial correlation in duty cycle values; regions in close proximity often exhibit similar duty cycles. Consequently, to effectively interpolate these duty cycle values, we employ a nonparametric approach that accounts for this spatial correlation, specifically inverse distance weighted (IDW) interpolation.

The IDW interpolation is a deterministic method used for estimating an unknown value, denoted as $\hat{s}_0$, at an unmeasured location x_0. The estimate is computed using a weighted average of known values, s_i, from measured points represented as x_i for $i \in \{1, 2, \ldots, m\}$ [10], [11],

$$\hat{s}_0 = \sum_{i=1}^{m} z_i s_i. \qquad (7)$$

The interpolation weight, z_i, is computed in a way that is greater for closer measured points,

$$z_i = \frac{\frac{1}{d(x_i, x_0)}}{\sum_{j=1}^{m} \frac{1}{d(x_j, x_0)}}, \qquad (8)$$

where $d(.)$ is the Euclidean distance function. This method assumes that closer points have more influence on the estimated value at the unknown location.

D. Long Term Signal Variation

Unlike signal variation at a shorter time frame due to shadow or multipath fading, the signal variation in a longer period such as months is difficult to associate with a single parameterized distribution. The long-term variation in signal strength is due to a diverse array of factors including short-term fading, daily transmit power variations depending on the time of the day (e.g. due to sleeping cells), and longer-term power variations depending on the month or season (e.g. due to seasonal change in population density in certain regions).

Once the average occupancy power levels are estimated throughout the map, the relation between power and signal variation is leveraged to estimate signal variation in the region. A nonlinear regression model is fit to the combined observations to estimate signal variation as a function of power as illustrated in Figure 6. The proxels at which signal strength estimate lies in an interval that corresponds to a negative variation estimate, the signal variation is calculated through IDW interpolation described by Equation (7).

E. Confidence Level

The confidence level is an important metric determining how reliable the other predicted RF monitoring metrics such as average occupancy power and signal variation are at a given proxel. The formulation of the confidence level depends on three factors [1]: the distance from the current proxel to the closest observed proxel, denoted as d_p; the maximum distance allowed from an observed proxel, represented by d_{max}; and the accuracy of the propagation model if it is used. Since the power, duty cycle, and variation calculations are largely based on observations, we formulate the confidence level purely as a function of distance factors d_p, and d_{max},
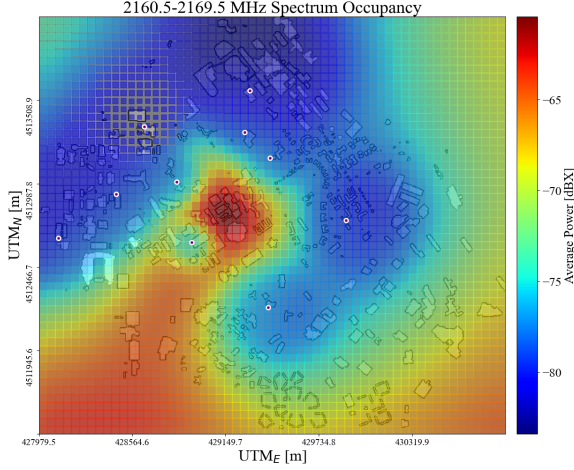
(a) 2160.5-2169.5 MHz
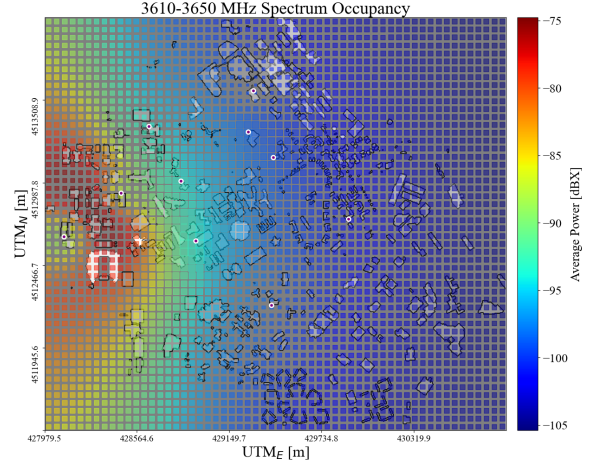
(b) 3610-3650 MHz

Fig. 4: Spatial spectrum occupancy, University of Utah campus

$$\gamma_p = (1 - \beta(d_p))e^{-\alpha d_p},$$
$$\beta(d_p) = \min(\frac{d_p}{d_{max}}, 1), \tag{9}$$

where γ_p is the confidence level at a proxel p, and α is an external parameter determining the steepness of the exponential function.

III. RESULTS & ANALYSIS

The methodology described for average occupancy power extrapolation is implemented on 2160.5-2169.5, 3470-3510 (C), and 3610-3650 MHz (CBRS) frequency bands. The process of estimating power in the CBRS band is depicted in Figure 3. To minimize the L_2 norm of the error vector defined in Equation (2), we utilize the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [12], with the estimated optimal transmit powers presented in Figure 3a. The calculated transmit power is then integrated into Equation (**??**), and the likelihood function is calculated facilitating the creation of a 2D probability mass function for transmitter locations, as illustrated in Figure 3b. The final step involves estimating the signal strength at every potential receiver location. This is achieved by employing a simple log-distance path loss model, which incorporates the previously estimated transmit powers. We then calculate the overall signal strength by multiplying the estimates of the path loss model with probabilities and summing across all proxels as outlined by Equation (6) and show the result in Figure 3c. The evaluation parameters used in this process are listed in Table I along with other parameters.

The estimation statistics are then calculated for each frequency band and listed in Table II. Results show that the variance of the average occupancy power across different sensors can be captured by up to 82%.

Once the occupancy power estimation is completed, the duty cycle estimates are computed using the IDW technique, and

TABLE I: Evaluation parameters.

Parameters	Values	Descriptions
r_p	5	Proxel width & height [m]
δ_c	400	Correlation coefficient [m]
σ	4.5	Standard deviation [dB]
m	10	Number of sensors
n_p	2	Path loss exponent
w	2.9	Overall map width [km]
h	2.6	Overall map height [km]
d_{max}	1	Max. distance from an observed proxel [km]
α	0.01	Steepness parameter [1/m]

TABLE II: Statistics

Freq. Band [MHz]	Init. Var. [dB²]	MSE [dB²]	Reduction %
2160.5-2169.5	18.19	9.50	47.8
3470-3510	25.92	7.56	70.8
3610-3650	41.37	7.35	82.2

the resulting estimates of power and duty cycle are combined into a single plot and illustrated in Figure 4. In the pixelated illustrations, colors represent the average occupancy power whose range is displayed on the side of the figure. Sizes of the square pixels on the other hand are proportional to the duty cycle estimated at that location. It is easily observed then that the 2160.5-2169.5 MHz signal is more frequently detected throughout the region than the CBRS band signal. It is also observed that regions on the map with lower duty cycles frequently align with a line of densely packed buildings that are situated between an area where the signal strength is high, likely near a transmitter, and the monitoring station in the low-duty cycle region. Also, it is often seen that these monitors in low-duty cycle areas are positioned immediately behind a building when observed from the high-signal power region potentially leading to large shadowing losses, which could result in frequent signal losses at the monitoring station. The figures also illustrate the areas where the signal strength is strong or weak, giving insights into the probable locations
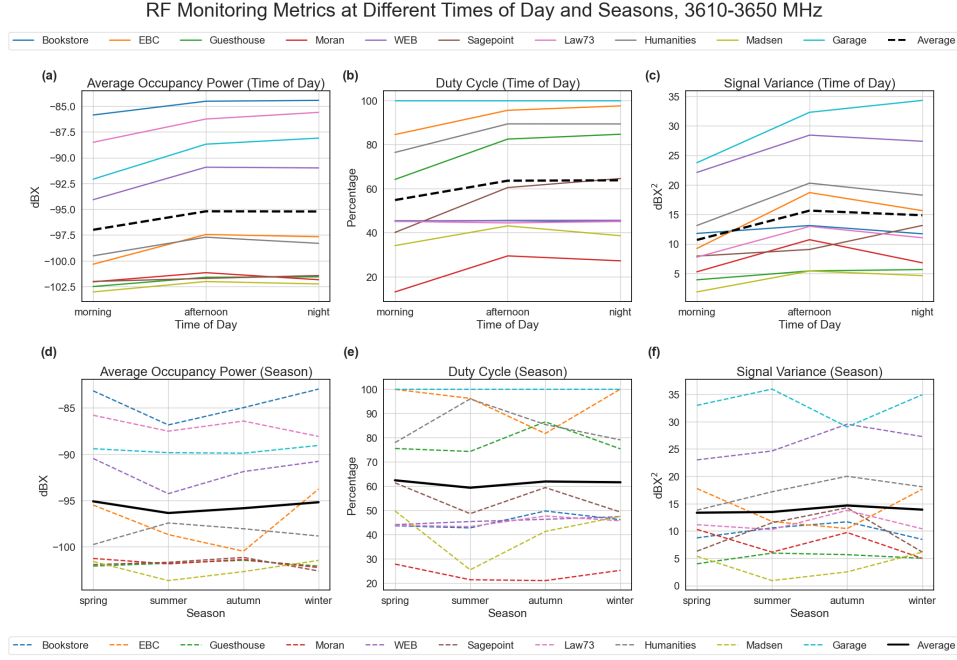
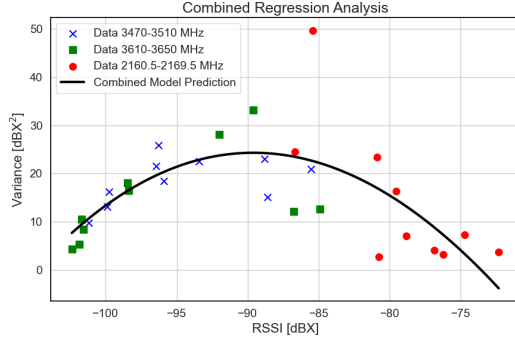Fig. 5: RF monitoring metrics at different times of day and seasons, 3610-3650 MHz

Fig. 6: Nonlinear regression

of transmitters, the extent of signal coverage, and the analysis of interference in certain parts of the region.

The trends of RF monitoring metrics over different times of day and seasons are illustrated in Figure 5 where morning is defined to be 4 am to noon, afternoon is noon to 8 pm, and night is 8 pm to 4 am. The average trend shows that the power, duty cycle, and variance are at lowest in the morning aligning with the sleeping cell concept. Similarly, most of the metrics drop to their lowest in the summer as expected since the population density on the campus is lower during the summer.

It is also observed that the metrics have similar trends as a function of time of day, and season. Accordingly, the correlations between pairs of metrics are analyzed and reported in Table III. The relationship between average occupancy power and variance is further analyzed as shown in Figure

6. The lower variance observed at the low-power regime is potentially due to "left-censored data" because powers lower than the threshold are not included in variance calculation [13]. The lower variance at the high-power regime is probably because most high-RSSI samples are LOS [14] and therefore less affected by fading [15]. The nonlinear regression model fit to the combined dataset is then used to predict signal variation throughout the region.

Finally, the confidence levels on the map, calculated by proximity to the nearest sensor, are merged with variation estimates in a single plot as shown in Figure 7. In these pixelated illustrations, colors indicate the signal variation, with the range displayed beside the figure. Pixel sizes correspond to the confidence level at that location. We note significant drops in confidence across regions where sensor proximity is limited. Signal strength variation also shows that variation is most pronounced when the average occupancy power is moderately strong. These estimates are useful for worst-case and best-case interference and coverage analyses.

TABLE III: Correlation coefficients between different metrics.

Frequency Band [MHz]	Pairs of Metrics	Pearson Correlation Coefficient
2160.5-2169.5	Variance-Mean Power	−0.76
	Variance-Duty Cycle	0.16
	Duty Cycle-Mean Power	0.00
3470-3510	Variance-Mean Power	0.46
	Variance-Duty Cycle	0.38
	Duty Cycle-Mean Power	−0.17
3610-3650	Variance-Mean Power	0.51
	Variance-Duty Cycle	0.48
	Duty Cycle-Mean Power	0.01
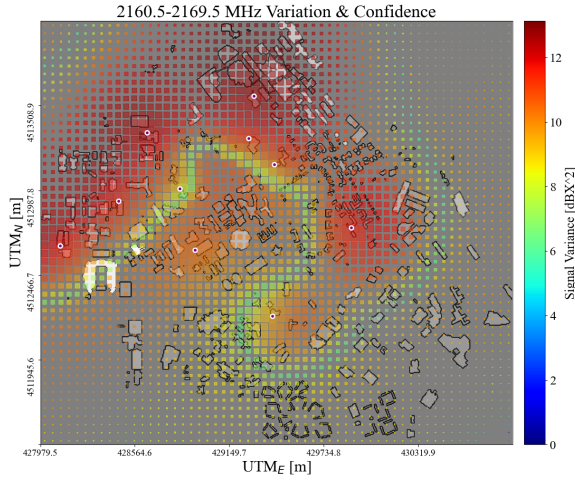
(a) 2160.5-2169.5 MHz
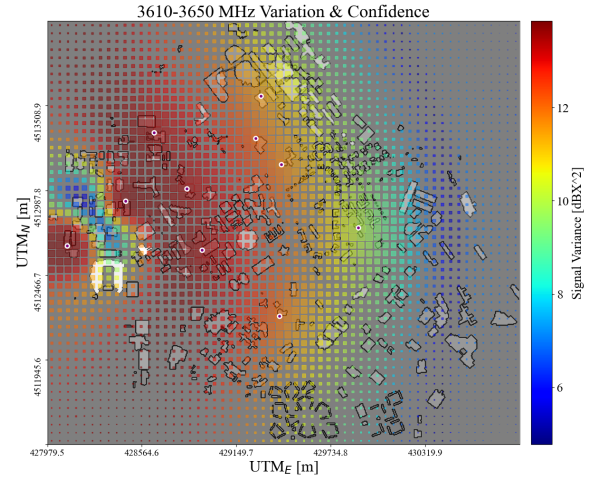
(b) 3610-3650 MHz

Fig. 7: Signal variation and confidence, University of Utah

IV. Conclusion

This paper addresses some of the challenges associated with the visualization and interpretation of spectrum occupancy in the dynamic domain of the radio frequency spectrum. The proposed solution makes use of a digital spectrum twin to overcome the complexities of the invisible and transient RF environment. Our contributions include a novel likelihood estimation technique that uses data from fixed spectrum monitors within the POWDER testbed to extrapolate spatially sparse average occupancy power observations. We also introduce a method for extrapolating signal variation observations, along with georeferenced visualizations that merge power & duty cycle in a single plot and signal variation & confidence levels in another. These visualization methods offer insights for identifying underutilized frequencies, future spectrum allocations, and broadcast coverage and interference analysis. The paper delivers an analysis of daily and seasonal spectrum usage trends, paving the way for automated and efficient RF spectrum management.

References

[1] S. Tadik, K. M. Graves, M. A. Varner, C. R. Anderson, D. Johnson, S. K. Kasera, N. Patwari, J. Van der Merwe, and G. D. Durgin, "Digital spectrum twins for enhanced spectrum sharing and other radio applications," *IEEE Journal of Radio Frequency Identification*, pp. 1–1, 2023.

[2] G. D. Durgin, M. A. Varner, N. Patwari, S. K. Kasera, and J. Van der Merwe, "Digital spectrum twinning for next-generation spectrum management and metering," in *2022 IEEE 2nd International Conference on Digital Twins and Parallel Intelligence (DTPI)*, 2022, pp. 1–6.

[3] G. D. Durgin, M. A. Varner, M. A. Weitnauer, J. Cressler, M. M. Tentzeris, A. Zajic, S. Zeinolabedinzadeh, R. Zekavat, K. Pahlavan, U. Guler, and K. V. d. Merwe, "Digital spectrum twinning and the role of rfid and backscatter communications in spectral sensing," in *2021 IEEE International Conference on RFID Technology and Applications (RFID-TA)*, 2021, pp. 89–92.

[4] J. Breen, A. Buffmire, J. Duerig, K. Dutt, E. Eide, A. Ghosh, M. Hibler, D. Johnson, S. K. Kasera, E. Lewis, D. Maas, C. Martin, A. Orange, N. Patwari, D. Reading, R. Ricci, D. Schurig, L. B. Stoller, A. Todd, J. Van der Merwe, N. Viswanathan, K. Webb, and G. Wong, "POWDER: Platform for open wireless data-driven experimental research," *Computer Networks*, vol. 197, Oct. 2021, https://doi.org/10.1016/j.comnet.2021.108281. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128621003017

[5] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. USA: Prentice Hall PTR, 2001.

[6] N. Patwari, A. Hero, M. Perkins, N. Correal, and R. O'Dea, "Relative location estimation in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2137–2148, 2003.

[7] A. Coulson, A. Williamson, and R. Vaughan, "A statistical basis for lognormal shadowing effects in multipath fading channels," *IEEE Transactions on Communications*, vol. 46, no. 4, pp. 494–502, 1998.

[8] N. Patwari and P. Agrawal, "Effects of correlated shadowing: Connectivity, localization, and rf tomography," in *2008 International Conference on Information Processing in Sensor Networks (ipsn 2008)*, 2008, pp. 82–93.

[9] P. Agrawal and N. Patwari, "Correlated link shadow fading in multi-hop wireless networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 8, pp. 4024–4036, 2009.

[10] A. E. C. Redondi, "Radio map interpolation using graph signal processing," *IEEE Communications Letters*, vol. 22, no. 1, pp. 153–156, 2018.

[11] Q. Liang, S. Nittel, J. C. Whittier, and S. de Bruin, "Real-time inverse distance weighting interpolation for streaming sensor data," *Transactions in GIS*, vol. 22, no. 5, pp. 1179–1204, 2018. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/tgis.12458

[12] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York, NY, USA: Springer, 2006.

[13] M. C. Newman, P. M. Dixon, B. B. Looney, and J. E. Pinder III, "Estimating mean and variance for environmental samples with below detection limit observations 1," *JAWRA Journal of the American Water Resources Association*, vol. 25, no. 4, pp. 905–916, 1989.

[14] T. S. Rappaport, S. Sun, and M. Shafi, "Investigation and comparison of 3gpp and NYUSIM channel models for 5g wireless communications," *CoRR*, vol. abs/1707.00291, 2017. [Online]. Available: http://arxiv.org/abs/1707.00291

[15] H. Poddar, T. Yoshimura, M. Pagin, T. Rappaport, A. Ishii, and M. Zorzi, "Ns-3 implementation of sub-terahertz and millimeter wave drop-based nyu channel model (nyusim)," in *Proceedings of the 2023 Workshop on Ns-3*, ser. WNS3 '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 19–27. [Online]. Available: https://doi.org/10.1145/3592149.3592155