

Assessing Interference with Regression Analysis Techniques

Jonathan E. Swindell

*Department of Electrical and
Computer Engineering
Baylor University
Waco, TX*

Jonathan_Swindell1@baylor.edu

Carson Slater

*Department of Statistical Science
Baylor University
Waco, TX*

Carson_Slater1@baylor.edu

Samuel Hussey

*Department of Electrical and
Computer Engineering
Baylor University
Waco, TX*

Samuel_Hussey1@baylor.edu

Charles Baylis

*Department of Electrical and Computer Engineering
Baylor University
Waco, TX*

Charles_Baylis@baylor.edu

Robert J. Marks II

*Department of Electrical and Computer Engineering
Baylor University
Waco, TX*

Robert_Marks@baylor.edu

Abstract—Reliably predicting aggregate interference is critical for deploying next-generation Dynamic Spectrum Access (DSA) systems. Next-generation DSA systems must operate in real-time. In current large-scale networks, aggregate interference prediction takes hours to complete. This work proposes regression analysis as a method for predicting interference. The model proposed in this work performs comparably to other machine learning approaches. Regression techniques can be fit quickly, unlike many machine learning methods which require long training. This rapid model training enables applications in dynamic spectrum environments. Our model achieved a mean absolute percentage error (MAPE) of 5.11% interference prediction with an 80/20 training validation split based on simulated data with added thermal noise density to approximate the noise of the transmission channel.

Keywords—aggregate interference, spectrum management, dynamic spectrum access, regression analysis, linear regression, generalized additive models

I. INTRODUCTION

Fifth-generation (5G) mobile broadband has enabled improvements in data rates, latencies, and coverage over previous generations at the cost of an increasingly crowded spectrum. High bandwidth demands required to support these developments have presented challenges to wireless services in adjacent frequency bands. Where separation between bands previously provided adequate protection from out-of-band (OOB) emissions, recent 5G allocations have resulted in operating frequencies close to and interleaved among existing services sensitive to interference. Despite 5G standards and policy measures designed to mitigate harmful interference scenarios, concerns remain in certain circumstances, evident by disagreements between compatibility studies on necessary protection levels for passive radiometers operating in the 23.6 – 24.0 GHz Earth Exploration-Satellite Service (EESS) [1]. This specific instance of concern arises from the auction of 24 GHz

spectrum to 5G wireless services, believed to present a notable risk to the extremely sensitive equipment operating in the EESS band. Designed to sense trace atmospheric microwave emissions, these radiometers are unable to adjust frequency or tolerance. This enhanced risk of harmful interference from OOB 5G emissions is thought to potentially lead to delayed and less accurate weather forecasts [2].

To address these challenges, new paradigms in spectrum policy have been developed, a notable example being Dynamic Spectrum Access (DSA). These DSA systems can coordinate frequency assignment independently of prior allocations, utilizing spectrum opportunistically to potentially reduce harmful interference, especially in services like the EESS. However, effective prediction and prevention necessitates the use of transmitter and receiver operating characteristics (e.g., location, power, radiation pattern, and gain). Acquiring this information with sufficiently low error and selecting an appropriate propagation model for specific environments is challenging. In complex propagation environments, especially environments where passive scientific services operate sensitive receivers, minor errors in propagation modeling, coupled with noise and multipath effects, can make deterministic interference calculations difficult. The complexity is further amplified when considering multiple sources and simultaneous transmissions.

II. ALTERNATIVE APPROACHES

As an alternative to commonly used statistical and deterministic approaches, this work considers linear regression (LR) and generalized additive models (GAM), a superset of linear regression analysis, as methods for predicting aggregate interference. More traditional, widely used statistical approaches are based on recommendations by the International Telecommunication Union (ITU) [3]. Additional statistical approaches exist that deviate from the ITU's recommendations. These models are less general and consider specific use cases,

but some can achieve higher-accuracy predictions more quickly. Kusaladharma and Vijayandran simplify existing statistical methods to predict interference moments generating function by considering only finite area networks [4, 5]. Peng simplifies aggregate interference prediction for deep space earth stations from high-density fixed services (HDFS) emitters by using geometrically partitioned regions to model the correlation of interferences [6]. Bhattacharai discloses that the log-normal distribution is an accurate estimate of aggregate interference for a fixed number of transmitters distributed in a network equidistant from one another [7,8].

Another statistical approach is taken by Ghasemi and Sousa, who developed a statistical model describing aggregate interference by considering the contributions of transmission power, path loss, channel fading, and receiver sensitivity [9].

Machine learning (ML) techniques show promise in aggregate interference detection. Saija examines numerous ML approaches to determine channel state information in 5G systems by predicting the signal-to-noise ratio and his results outperform traditional approaches [10]. The successful application of ML here shows promise for the application of ML in similar problems.

Padilla developed a non-linear autoregressive neural network (NARNN) that predicts interference and has been shown to enable more efficient resource allocation [11]. Zhao demonstrates a neural network to predict aggregate interference trained on location, transmit power, and path loss [12]. Hussey presents a method to detect aggregate interference using fuzzy inference systems providing intuition and fast operation in field deployment [13].

These developments are significant in the journey to providing viable methods to predict aggregate interference, but the authors are not aware of any of the above techniques having been deployed at scale in real time. Real-time operation is needed to enable the next generation of reconfigurable spectrum-sharing systems. Laborious coexistence calculations that limit dynamic allocation speed are clearly a hindrance for scenarios and services with real-time needs, such as 5G/radiometer coexistence at 24 GHz. Another specific example of this need can be found in the Citizens Broadband Radio Service (CBRS) Spectrum Access System (SAS) aggregate interference prediction methodology, where calculations often delay assignments up to 24 hours [14].

III. DATA GENERATION

A synthetic dataset comprised of transmit power spectral densities (PSDs), distances between devices, and subsequent PSDs after propagation losses were generated by simulating a simple wireless network on a spectral brokering platform based on the work in [15]. In these simulations, Spectrum Consumption Models (SCMs) were used to articulate various aspects of spectrum use. For our purposes, SCMs are data structures in which transmitter models capture the origin, frequency, strength, and direction of emissions, while receiver models define interference thresholds and frequencies of reception [16]. The dataset was captured by varying the location of a radiometer model across the network's 1 km² area while a set of three 5G base stations remained fixed in place. The center

operating frequency of the radiometer was chosen to be 23.84 GHz with a 200 MHz channel width. For simplicity, frequencies of the out-of-band transmissions from the adjacent 5G-NR n258 band were modeled using the same center and bandwidth.

To determine the extent of interference at each location, each emission is scaled by free space path loss in the appropriate direction. The final dataset consists of the out-of-band transmission PSDs (ranging from -70 to -40 dBW/200 MHz), distances between base stations and radiometer (ranging approximately from 50 to 1650 meters), and the interfering powers. Fig. 1 illustrates an instance in which emissions of similar powers experiencing path losses are incident upon the radiometer location. This scenario was simulated approximately 20000 times with varying radiometer locations, each location providing three data points to be used in the analysis.

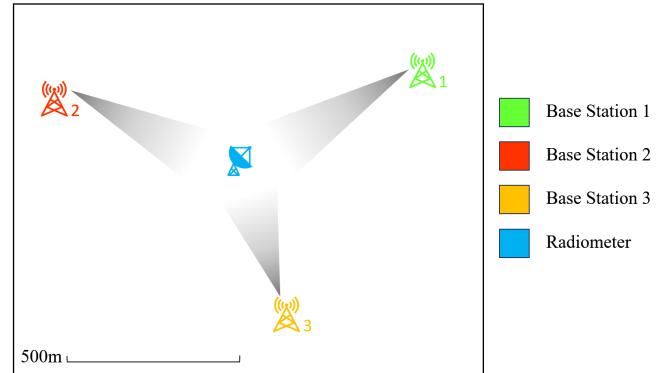


Fig. 1. Device placement in a network area of approximately 1 km²

IV. REGRESSION ANALYSIS

Although ML approaches such as neural networks can prove to be extremely accurate without requiring many statistical assumptions, they are computationally expensive. Machine learning generally exploits correlations within data to develop accurate predictions, but can be hard to interpret. This is why many ML models are referred to as 'black box models.' Regression models, although rudimentary, could prove useful as they are interpretable and computationally inexpensive. If acceptable results are achieved, the simplicity and low computational cost of regression models could make them an ideal tool for real-time prediction of dynamic aggregate interference. The analysis presented here was performed on simulated device pairs. The benefits of these techniques support large-scale real-time aggregate interference prediction.

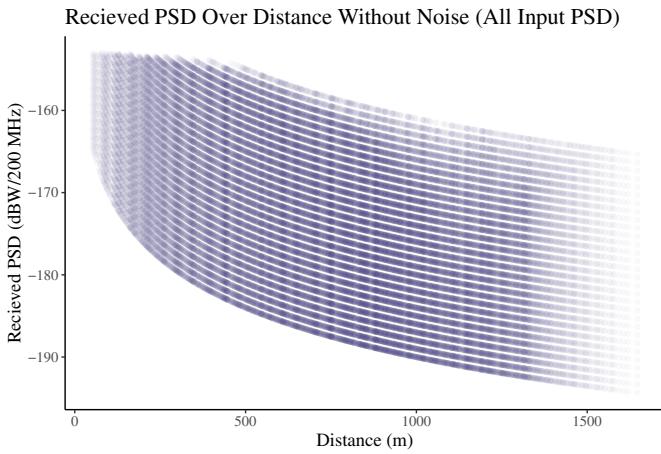
Before using a regression model to predict output PSD, it must be acknowledged that a model is only as reliable as its training data and assumptions. Because the training data is generated from simulation, the data used to fit the model is deterministic, as opposed to stochastic. To better model real-world propagation, noise, or jitter, has been added to the training data. Jitter is a viable tool to improve a model's generality when small uniform noise perturbs a system [17]. Thermal noise density has been chosen as an estimate of noise because our model considers PSD. Thermal noise power is given as:

$$P_N = kTB \quad (1)$$

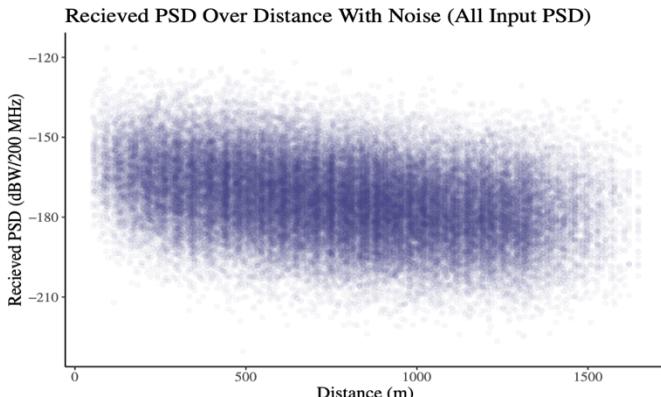
Where k is the Boltzmann's constant, T is the temperature, and B is the bandwidth. Thermal noise density is obtained by disregarding bandwidth:

$$D_N = kT \quad (2)$$

The thermal noise density at room temperature is approximately -174 dBm/Hz [18]. In units of dBW/200 MHz gives a thermal noise density of -120.99 dBW/200 MHz. By assuming that this value is the variance, random Gaussian noise is added to the received PSD. The added noise is distributed normally with a mean of $\mu = 0$ dBW/200MHz and a standard deviation of approximately $\sigma = 10.99$ dBW/200MHz. A Gaussian probability model was selected because of the possibility of both constructive and destructive interference, with the jitter being centered at zero. It is expected that fitting a model on a dataset with simulated noise increases the robustness of the model and raises the likelihood of success in extrapolating the model to real-world applications in future studies.



Raw simulated data plotted before any added noise without the distance variable transformed.



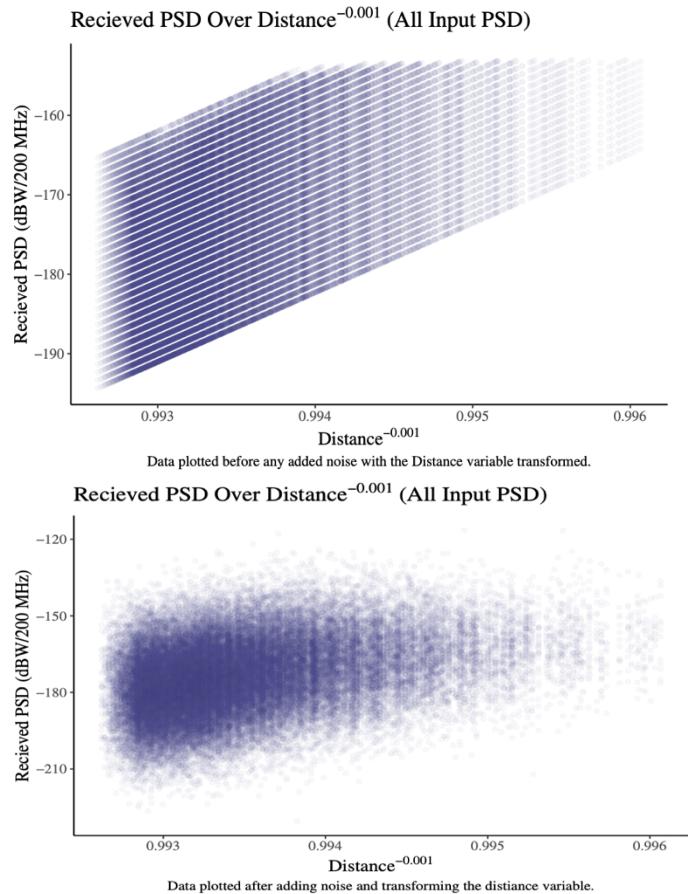
Raw simulated data plotted with simulated noise without the distance variable transformed.

Fig. 2. (a) Synthetic data with input PSD ranging from -41 to -70, before any transformation, (b) synthetic data with added jitter. Plots generated with [19].

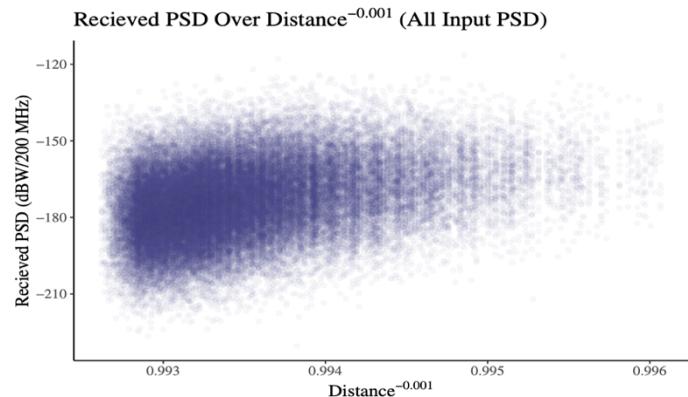
In the model, two predictors are proposed for the PSD received by the radiometer: the transmitted PSD (dBW/200MHz) and the distance between the radiometer to the transmitter (meters). A regression model is sought that can predict the PSD received by the radiometer at a given distance and transmitted power spectral density. Linear regression models are beneficial for identifying the strength and magnitude of the associations between predictors and outcomes.

Traditionally, multiple linear regression models perform best when there is a strictly linear relationship, with no indication of higher-order terms necessary to model the phenomena. In our case, we observe a nonlinear relationship between distance and received PSD (Fig. 2(a)) that is not sufficiently captured by a higher-order polynomial. Note how Fig. 2(a) resembles a family of curves, resulting from the unique transmitted PSD levels. Hence, it is attempted to transform the data to resemble a linear model through an arbitrary mapping of the distance feature to generate a linear relationship. This transformation is applied and visualized in Fig. 3. This linear relationship can then be applied to a regression model. This type of method is broadly known as a generalized additive model (GAM).

A GAM essentially contains model features that are functions of input variables, while still enabling it to be fit with coefficients like a multiple regression model.



Data plotted before any added noise with the Distance variable transformed.



Data plotted after adding noise and transforming the distance variable.

Fig. 3a. Synthetic data for the regression model with input PSD ranging from -41 to -70, without the added Gaussian noise. 3b Synthetic data under transformation with added jitter. Plots generated using [19].

The model's proposed distance feature mapping is

$$f(x) = x^{-0.001}, f: (\mathbb{R} \rightarrow \mathbb{R}), \quad (3)$$

as it was found to be the most reasonable and effective transformation for creating a linear relationship. One of the strengths of a GAM is that it allows the flexibility to fit a non-

linear relationship (f) to the distance variable. Fig. 2(a) shows that the relationship between distance and received PSD is nonlinear, and Fig. 3(a) shows that the relationship becomes linear through the proposed transformation. This transformation was chosen because after exploring many possible functions it best transforms the data to resemble a linear relationship. Fig. 3(b) also shows the simulated noise imposed on the data.

To discern the relationship between true distance and received PSD, one can take the value in the domain of f and apply the inverse mapping f^{-1} . To discern the relationship between the transmitted PSD and the received PSD, no inverse mapping is required, because no transformation is included in the model for that predictor. This predictor is a discrete variable with a linear relationship. Our proposed GAM is as follows:

$$\hat{R}_{PSD} = \hat{\beta}_0 + \hat{\beta}_1 T_{PSD} + \hat{\beta}_2 f(D) \quad (4)$$

Where T_{PSD} is the transmitted PSD, D is the distance, and R_{PSD} is the received PSD. Each coefficient ($\hat{\beta}_i$) serves as an estimated weighted ‘slope’ for the respective term [20]. The coefficients for the models with and without added noise are given in Table I and Table II. Table III gives the error metrics for the models. The model was fit using an 80/20 training testing data validation split, where 80% of the entire data set was used for training, while the other 20% was reserved for model evaluation. This precaution was taken to prevent overfitting.

Metrics chosen to gauge the model’s predictive power are MAPE (mean absolute percentage error), and RMSE (root mean square error), where RMSE is in units of the outcome of interest. These metrics gauge the relative absolute error and the magnitude of the error in the units of the outcome variable respectively. R^2 was included to show the percentage of the variability in the testing data that was explained by the model. It was used to determine whether or not a model was overfit. R^2 is most useful when there is a linear relationship between the outcome and predictors, which was not originally the case, but after transforming the distance variable Fig. 3(a), there exists a linear relationship between the distance and the received power spectral density.

TABLE I: GAM (No Noise)

Coefficient	Estimate	Std. Error	P-Value
(Intercept)	-8799.755	0.0146	0
<i>Input PSD</i>	1.000022	$1.05e^{-6}$	0
$Distance^{-0.001}$	8739.921	0.0147	0
$R^2: 1.00 MAPE: 0.0008\%$ [21]			

TABLE II: GAM (Noise)

Coefficient	Estimate	Std. Error	P-Value
(Intercept)	-8883.415	80.19	0
<i>Input PSD</i>	0.995144	0.005	0
$Distance^{-0.001}$	8823.817	80.73	0
$R^2: 0.401 MAPE: 5.11\%$ [21]			

TABLE III: COMPARING METRIC BETWEEN NOISE/NO NOISE MODEL

Metric	No Noise	Noise
MAPE	0.000834%	5.11%
RMSE	0.00197	11.0
R^2	1.00	0.401

Table II shows the model coefficients when noise is added to the GAM. With added jitter, more realistic model metrics were obtained, indicating the model was not overfit to synthetic data. Table III provides a comparison of the no-noise model and the model including noise. The no-noise model results in model coefficient estimates with a very small standard error, an extremely small mean absolute percentage error (MAPE), and also a perfect R^2 . Thus, these metrics indicate potential overfitting. The noise-added model demonstrates model coefficients with higher standard error, much greater (yet excellent) MAPE, and a R^2 of around 0.4. The coefficient estimates across models are similar, because the probability model underlying the noise is Gaussian, which is symmetric. Regardless, the GAM trained on noisy data still performs comparably with ML models.

Although most classical statistical methods will be outperformed by ML models, GAM’s are easier to interpret and also are far less computationally expensive, while achieving high performance and accuracy. 10-fold cross-validation was also used to avoid overfitting. For this model, this required under two seconds to fit fifty models on subsets of the training data [21]. In the analysis, 10-fold cross-validation was performed five times. Each trial featured an 80-20 training-testing split, and a model was fit and tested on the respective testing subset. Upon correctly specifying the model, output PSD can be predicted, given values of input PSD and distance from a radiometer.

To visualize the model performance, Fig. 4 features a subset of the data where the input PSD is -55 dBW/200MHz, and shows a point estimate for the model, as well as a 95% prediction interval. A prediction interval is different from a confidence interval: this is a 95% interval that is expected to contain the next future unobserved observation (i.e. the $n + 1^{th}$ observation).

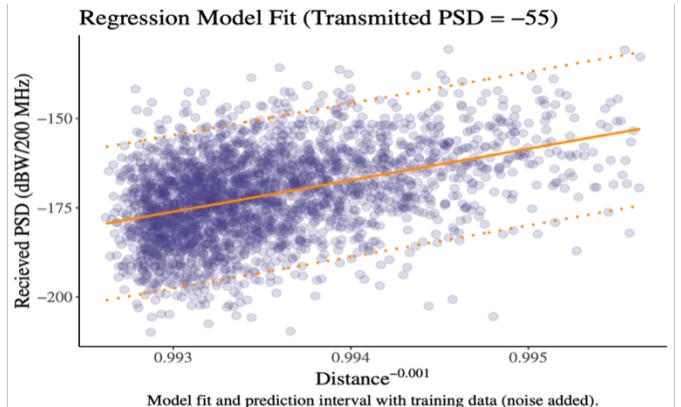


Fig. 4. Holding fixed the input PSD at -55 dBW/200MHz relationship between the transformed distance and the outcome, received PSD (dots) with added noise. The orange line is the predicted values of the model fitted on the noisy training data, and the dashed lines denotes the 95% prediction interval. Plots generated using [19].

Because the model features two input variables and one output variable, its prediction can be visualized in three dimensions. Fig. 5(a) and Fig. 5(b) show the control surfaces and the model prediction behavior. Fig. 5(a) shows the control surface before transforming the distance feature to obtain a linear relationship. The control surface matches the intuition of the system because either a decrease in distance or an increase in transmitted PSD will increase the interfering PSD at the receiver. Fig. 5(b) shows the control surface in the transformed space. As expected, this surface is linear.

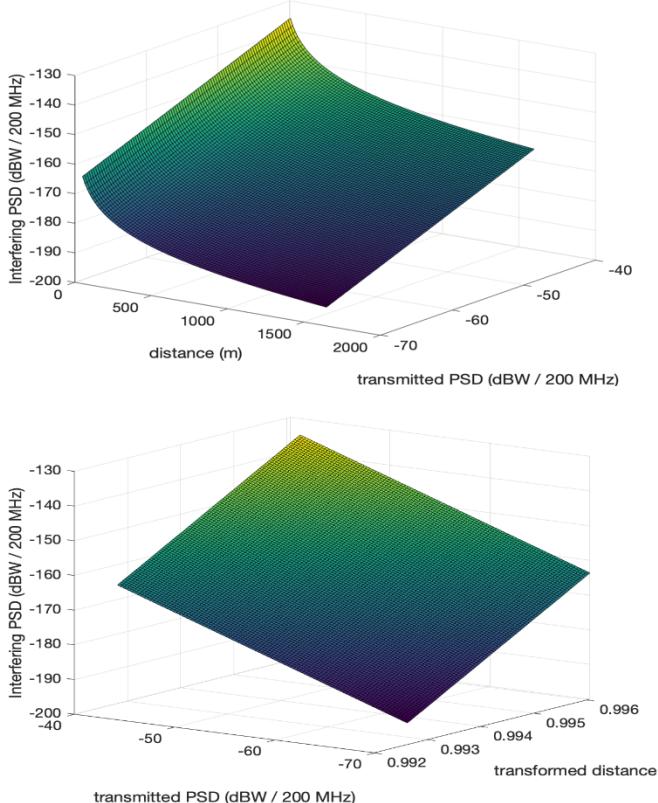


Fig. 5. (a) Control surface for the higher dimensional space of the model with the untransformed distance feature, (b) control surface for the higher dimensional space of the model with the transformed distance feature

Using the developed model, the PSD received by the radiometer can be predicted, given the transmitted PSD and the distance of the radiometer from the transmitter. Fig. 4 shows the model predictions across the distance feature holding fixed the transmitted PSD. The model has been trained over a distance range of 50 to 1600 meters and over a transmission PSD range of -70 to -41 dBW/200MHz and their corresponding interfering PSD. Our model should not be used to predict anything outside of that range. For example, given a distance of 500 meters and a transmission PSD of -60 dBW/200 MHz the model predicts an interfering PSD equal to -173.40 dBW/200 MHz.

While the study was conducted with synthetic data to which jitter has been applied, this study seems to indicate that this method will also be successful with noisy measured data. The next steps include testing this model on measurement data collected from actual transmitters, and validating and refining this model based on these results.

V. COMPARISON WITH ALTERNATIVE METHODS

The proposed method fares well compared to machine learning approaches. Table IV provides a comparison of relevant models used. For this study, all of the compared methods have been trained using simulated datasets.

TABLE IV: MODEL COMPARISON

Model	MAPE
GAM (no noise)	0.000834%
GAM (noise added)	5.11%
LR (no noise)	0.670%
LR (noise added)	5.16%
Fuzzy Inference System (no noise)	0.6755% [13]
Nonlinear Autoregressive Neural Network (noise added)	7.8% [11]

The GAM more accurately predicts interference with and without added noise, compared to other machine learning techniques trained on simulated data. In addition, this model can be fit, or trained, in milliseconds, whereas machine learning methods such as the Fuzzy Inference System can require hours to train [13]. As such, this linear regression method is well-suited for dynamic network environments that require frequent retuning. The same simulated data was used to train the fuzzy inference system, LR model, and the GAM. However, the simulated data used to train the Nonlinear Autoregressive Neural Network differs [11]. The LR results shown here are preliminary results found in this study. Our methods for obtaining the coefficient ($\hat{\beta}_i$) for LR are identical, but the transformation to linearize the relationship between our features was not applied.

VI. CONCLUSIONS

The applicability of LR and GAM modeling techniques to the prediction of interference has been thoroughly investigated. The GAM model performs comparably to previous machine-learning approaches. The proposed method enables rapid model training which can enable more real-time aggregate interference assessments in dynamic spectrum environments. A MAPE of 5.11% interference prediction with simulated data with added thermal noise density to approximate the noise of the transmission channel.

ACKNOWLEDGMENT

This work has been funded by the National Science Foundation (Grant No. 2030243).

REFERENCES

- [1] Summary of Studies Performed in TG 5/1 For Protection of Passive EESS in Band 23.6-24 GHz, ITU Doc CPM19-2/99-E (6 February 2019).
- [2] D. Lubar, D. Kunkee, and L. Cahsin, "Developing a Sustainable Spectrum Approach to Deliver 5G Services and Critical Weather Forecasts," in Center for Space Policy and Strategy, The Aerospace Corporation, January 2020.

- [3] "Characterization and assessment of aggregate interference to Earth exploration-satellite service (passive) sensor operations from multiple sources of man-made emissions", Recommendation ITU-R RS.1858 (01/2010).
- [4] S. Kusaladharma and C. Tellambura, "Aggregate Interference Analysis for Underlay Cognitive Radio Networks," in IEEE Wireless Communications Letters, vol. 1, no. 6, pp. 641-644, December 2012, doi: 10.1109/WCL.2012.091312.120600.
- [5] L. Vijayandran, P. Dharmawansa, T. Ekman and C. Tellambura, "Analysis of Aggregate Interference and Primary System Performance in Finite Area Cognitive Radio Networks," in *IEEE Transactions on Communications*, vol. 60, no. 7, pp. 1811-1822, July 2012, doi: 10.1109/TCOMM.2012.051412.100739.
- [6] Peng, T., Kinman, P., Kayalar, S., and Ho, C., "Estimating the Aggregate Interference from High-Density Fixed Service Emitters to Deep-Space Earth Stations", Interplanetary Network Progress Report, vol. 42-179, pp. 1-27, 2009.
- [7] S. Bhattacharai, A. Ullah, J. -M. J. Park, J. H. Reed, D. Gurney and B. Gao, "Defining incumbent protection zones on the fly: Dynamic boundaries for spectrum sharing," *2015 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Stockholm, Sweden, 2015, pp. 251-262, doi: 10.1109/DySPAN.2015.7343908.
- [8] S. Bhattacharai, J. -M. Jerry Park, W. Lehr and B. Gao, "TESSO: An analytical tool for characterizing aggregate interference and enabling spatial spectrum sharing," *2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Baltimore, MD, USA, 2017, pp. 1-10, doi: 10.1109/DySPAN.2017.7920793.
- [9] A. Ghasemi and E. S. Sousa, "Interference Aggregation in Spectrum-Sensing Cognitive Wireless Networks," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 41-56, Feb. 2008, doi: 10.1109/JSTSP.2007.914897.
- [10] K. Saija, S. Nethi, S. Chaudhuri and R. M. Karthik, "A Machine Learning Approach for SNR Prediction in 5G Systems," *2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*, Goa, India, 2019, pp. 1-6, doi: 10.1109/ANTS47819.2019.9118097.
- [11] C. Padilla, R. Hashemi, N. H. Mahmood and M. Latva-Aho, "A Nonlinear Autoregressive Neural Network for Interference Prediction and Resource Allocation in URLLC Scenarios," *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju Island, Korea, Republic of, 2021, pp. 184-189, doi: 10.1109/ICTC52510.2021.9620845.
- [12] Y. Zhao, L. Shi, X. Guo and C. Sun, "Aggregate Interference Prediction Based on Back-Propagation Neural Network," *2018 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, Seoul, Korea (South), 2018, pp. 1-5, doi: 10.1109/DySPAN.2018.8610415.
- [13] S. Hussey, J. Swindell, A. Goad, et al, "Assessing Aggregate Interference with Mamdani Fuzzy Inference Systems", *DySPAN*, 2024 (Submitted)
- [14] FCC Technical Advisory Council, "Recommendations to the Federal Communications Commission Based on Lessons Learned from CBRS," December, 2022. [Online]. Available: https://www.fcc.gov/sites/default/files/recommendations_to_the_federal.communications_commission_based_on_lessons_learned_from_cbrs.pdf
- [15] S.A. Seguin, A. Goad, C. Baylis, R.J. Marks, "Spectrum Sharing Brokers for Active and Passive Devices," *2022 IEEE International Symposium on Electromagnetic Compatibility and Signal/Power Integrity*, Spokane, Washington, August 2022.
- [16] IEEE, "IEEE 1900.5.2-2017 - IEEE standard for method for modeling spectrum consumption," 2017.
- [17] R. Reed, R. J. Marks and S. Oh, "Similarities of error regularization, sigmoid gain scaling, target smoothing, and training with jitter," in *IEEE Transactions on Neural Networks*, vol. 6, no. 3, pp. 529-538, May 1995, doi: 10.1109/72.377960.
- [18] G. Gonzalez, *Microwave transistor amplifiers* (2nd ed.): analysis and design. USA: Prentice-Hall, Inc., 1996.
- [19] Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org>.
- [20] Kuhn et al., (2020), Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. <https://www.tidymodels.org>
- [21] Bogetoft P, Otto L (2022). *Benchmarking with DEA and SFA*. R package version 0.31.