Ryerson University
Data Science & Analytics Graduate Studies

# 2018-2019 M.Sc. in Data Science and Analytics

# Radio Coverage Prediction in Urban Environment Using Regression-based Machine Learning Models

Sophie Roginsky
Supervisor: Dr. Ayşe Başar

## Objective

The objective of this research is to determine whether a Generalized Linear Model (GLM) is a viable solution for the prediction of radio signal strength in a fixed urban environment. An accurate propagation model that is linear in parameters can serve as a practical tool for coverage optimization applications in radio network planning.

## Background

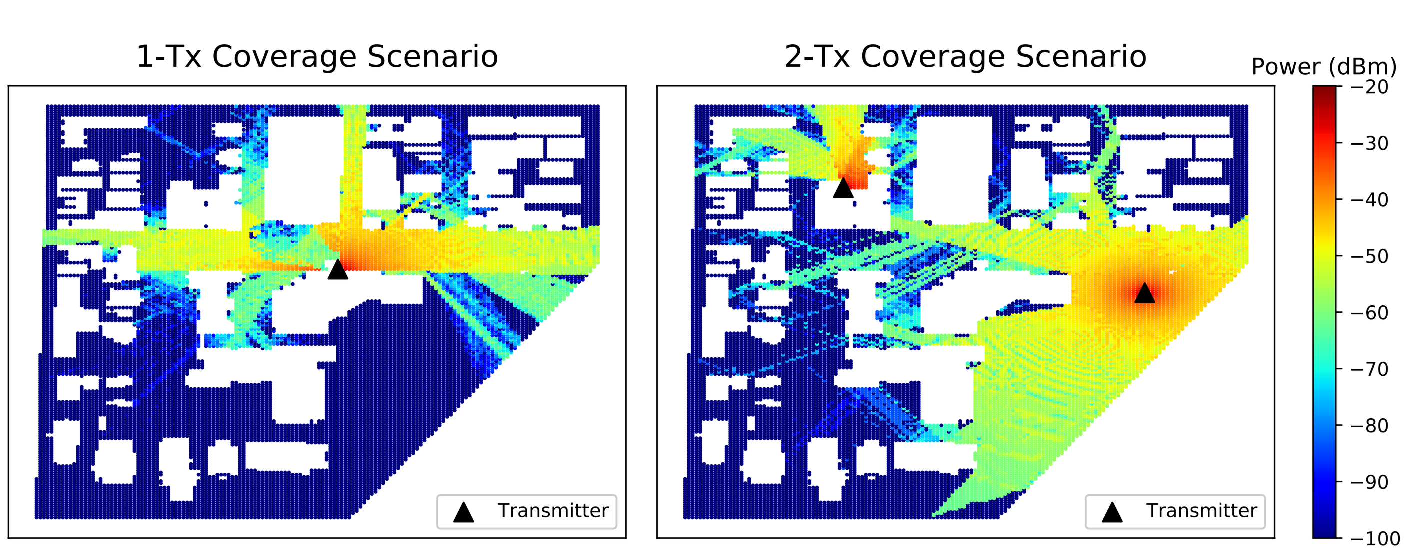
Planning a reliable radio network requires an accurate model of the radio signal strength in the intended environment. Propagation models are important tools for radio coverage estimation and the determination of optimal radio transmitter locations.

The prediction of radio signal strength can be viewed as a regression problem. The inputs are represented by information about the transmitter and environment, and the resulting power coverage represents the output to be calculated. The goal is to find a representative set of input features *(x)* and a function *f(x)* that can accurately estimate the power coverage, given a transmitter location. Machine learning methods are useful tools for solving regression problems, and have been effectively used for propagation prediction.

In existing literature, the most commonly used machine learning models for propagation prediction are Neural Networks and kernelized Support Vector Regressors. The drawback of these models is that they are non-linear in parameters and, therefore, cannot be optimized for coverage optimization applications. We propose a generalized linear model with propagation prediction accuracy comparable to non-linear regression models, with the advantage of being optimizable, interpretable and requiring less training time.

## Methodology

The data are a set of simulated radio propagation scenarios in a urban environment in downtown Ottawa, generated using a ray-tracing software. Each data instance represents a coverage heatmap over a transmitter location. We considered one and two transmitter scenarios in a fixed region, as depicted below.



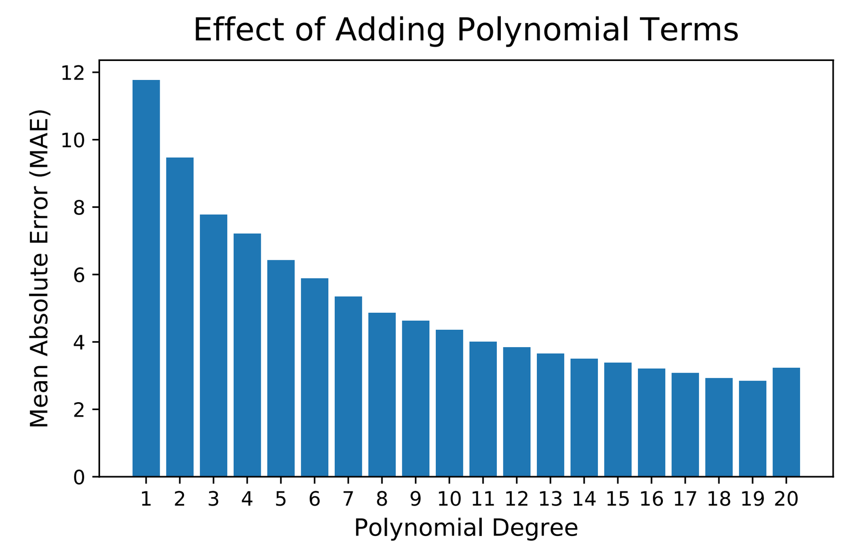

We began by simplifying the complexity of the problem by reducing the dimensionality of the grid. After building successful models on a coarse grid, the dimension of the grid was continually increased to approach the original grid size.

In order to increase the representational power of the features, we performed polynomial feature engineering. We selected the constructed features that offered the greatest improvement over the baseline performance as our final feature set.

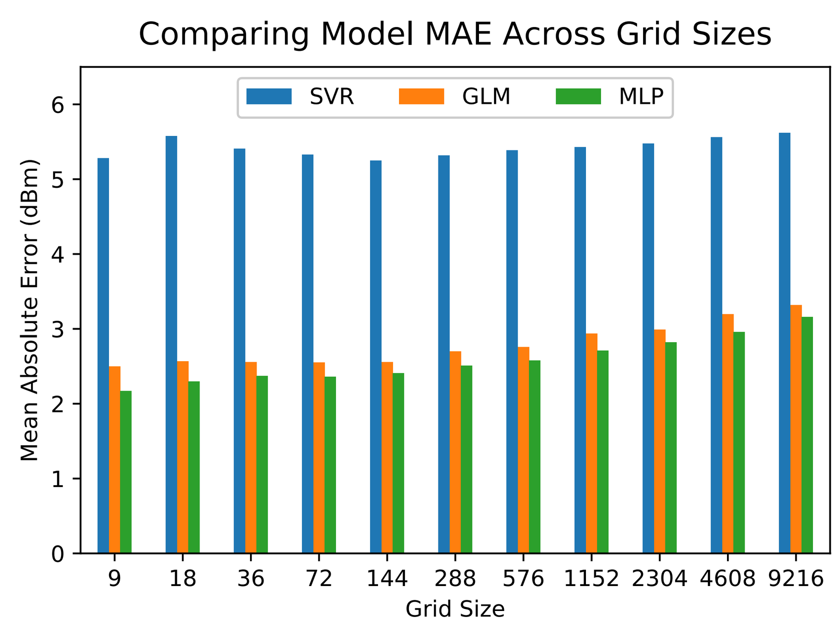
We trained a Generalized Linear Model (GLM) using the ordinary least squares method as our primary model. Once the model reached its optimal performance, we trained Multi-Layer Perceptron (MLP) and Support Vector Regressor (SVR) models to evaluate how well the linear model is performing in comparison to more complex, non-linear regression models.

## Results

Polynomial expansion, as shown on the right, proved to be an effective feature construction method for the GLM, offering a significant decrease in error with an increase of higher order polynomial terms.



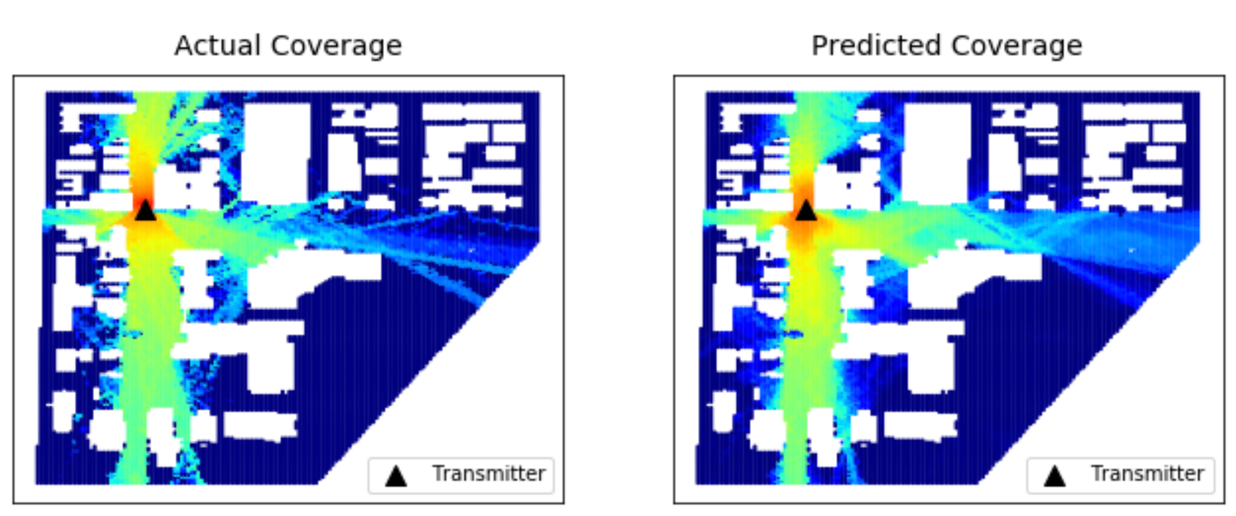

In the same manner, MLP and SVR models were trained and enhanced using polynomial feature construction methods for a final performance comparison between the three models. It is clear that SVR greatly underperformed the other two models. The GLM performance is comparable to the MLP, trailing only slightly behind in prediction accuracy, with dramatically lower training time.




| | SVR | GLM | MLP |
|---|---|---|---|
| MAE | 5.67 | 3.32 | 3.16 |
| $R^2$ | 0.69 | 0.89 | 0.9 |
| Time | 7 hrs | 16 secs | 6 hrs |

*Model performance on finest grid size*

Below is a sample visualization of the power coverage predicted by the GLM next to the actual power coverage generated by the ray-tracer. A visual inspection reveals that the linear model was able to learn the propagation patterns of a single transmitter, capturing the power and direction of the radio signal quite accurately.




## Conclusions

Regression-based models, such as Multi-Layer Perceptrons and Support Vector Regressors, have proven their success for the propagation prediction problem. Although offering high predictive accuracy, the disadvantage of these complex models is that their training procedure can be very computationally expensive. If used for the application of coverage optimization, another drawback is that these models they cannot be optimized in a straightforward fashion.

In this research, we proposed a simple Generalized Linear Regression Model that, when combined with appropriate feature construction methods, offers performance comparable to MLP and SVR models at a significantly lower training cost. Moreover, the GLM model is linear in its parameters, making it a practical option for coverage optimization problems.