



North Central Texas
Council of Governments

Integrating Transportation and Stormwater Infrastructure

Projected Maximum Potential Area of Bioretention Analysis

June 2026



*This study was funded by the Texas Water Development Board and
Texas Department of Transportation.*



1 TSI Projected Maximum Potential Area of Bioretention Analysis

Texas A&M AgriLife (Samantha Murray)

1.1 Overview

The purpose of this analysis was to determine the maximum potential area of bioretention for the TSI study areas projected up to 2070. This analysis is intended to predict the maximum potential bioretention area for 2045 and 2070 based on impervious land use in 2020. The land use model was developed by the NCTCOG and UT-Arlington using the best available local knowledge and identified three pilot watersheds with projected rapid urbanization and local or downstream impacts on flooding risks. The pilot areas are Bridgeport, Eagle Mountain, and Mary's Creek, and the urbanization is expected to be similar to that of the suburban areas in and around Dallas, TX. Therefore, the regression model developed for this analysis was based on the City of Dallas, excluding the industrial zones.

Bioretention areas are green stormwater infrastructure (GSI) used for stormwater management and are typically used in public and commercial zones such as parks, parking lots, roadway medians, and large sidewalks. The determination of the potential area of bioretention follows the guidance of the Green Stormwater Infrastructure for Urban Resilience: Opportunity Analysis for Dallas, Texas by Jack et al. (2022), additionally this source provided some of the data used in this analysis. Thus, candidate bioretention placement and their areas were estimated as 10% of parks and trails, 10% of parking lots, 35% of vegetated road medians, and 35% of non-residential sidewalks at least 8 feet (ft) wide.

1.2 Methodology approach

All analysis were completed using Esri ArcGIS Pro 3.5 software and Python 3

1.2.1 Data Collection

The geographic information systems (GIS) data was accessed through the City of Dallas GIS Data Hub and ParkServe by the Trust for Public Land. The Dallas city boundary, zoning, tax parcels, and sidewalks GIS vector data was developed by the City of Dallas. The Dallas vegetated medians and parking lots GIS vector data from Jack et al. (2022) which originally sourced this data from the Trust for Public Land, was accessed via the City of Dallas GIS Data Hub. The parks and trails GIS vector data acquired were produced by the Trust for

Public Land. The current (2020) and future (2045 and 2070) percent impervious land use raster data provided by the NCTCOG and UT-Arlington was at a spatial resolution of 90 meters (m).

- a. City of Dallas GIS Data Hub: city boundary, city zoning, tax parcels, and sidewalks.
- b. Jack et al. (2022) accessed via City of Dallas GIS Data Hub: vegetated medians and parking lots.
- c. Trust for Public Land: parks and trails
- d. NCTCOG and UTA: Percent Impervious Land 2020, 2045, 2070

1.2.2 Data Preprocessing

The parks, trails, parking lots, roadway vegetated medians, and sidewalks data was limited to the City of Dallas non-industrial zones via spatial analysis. The parks and trails layer were limited to those with public access based on the attribute data table. Sidewalks were limited to existing and greater than or equal to 8 ft in width based on the attributes and limited to non-residential areas according to tax parcels through spatial analysis. Buffers with a width based on defined sidewalk width were created around the line features and used to estimate the area. Parking lots and roadway medians were not further preprocessed in this analysis since they were used and preprocessed by the source for a similar analysis, it is noted that this data was limited to their studies focal subwatersheds of Dallas. These four layers were checked for overlaps and only existed between sidewalks and parks and trails as $> 1\%$, $\sim 5,200 \text{ ft}^2$, of the area and thus was considered to have no significant impact so it could be ignored.

These layers were used to calculate the total area and percentage of that area that could be converted to bioretention. Following the analysis by Jack et al. (2022), candidate bioretention placement and their areas were estimated as 10% of parks and trails, 10% of parking lots, 35% of vegetated road medians, and 35% of non-residential sidewalks at least 8 ft wide. These four polygon vector layers were then converted to raster data. Raster resolutions were determined based on the spatial size of each layer, parks/trails and parking lots at a 10 m resolution and vegetated medians and sidewalks at a 3 m. Each raster was aggregated to match the cell size of 90 m and calculated the percentage area of maximum potential bioretention (Figure 1).

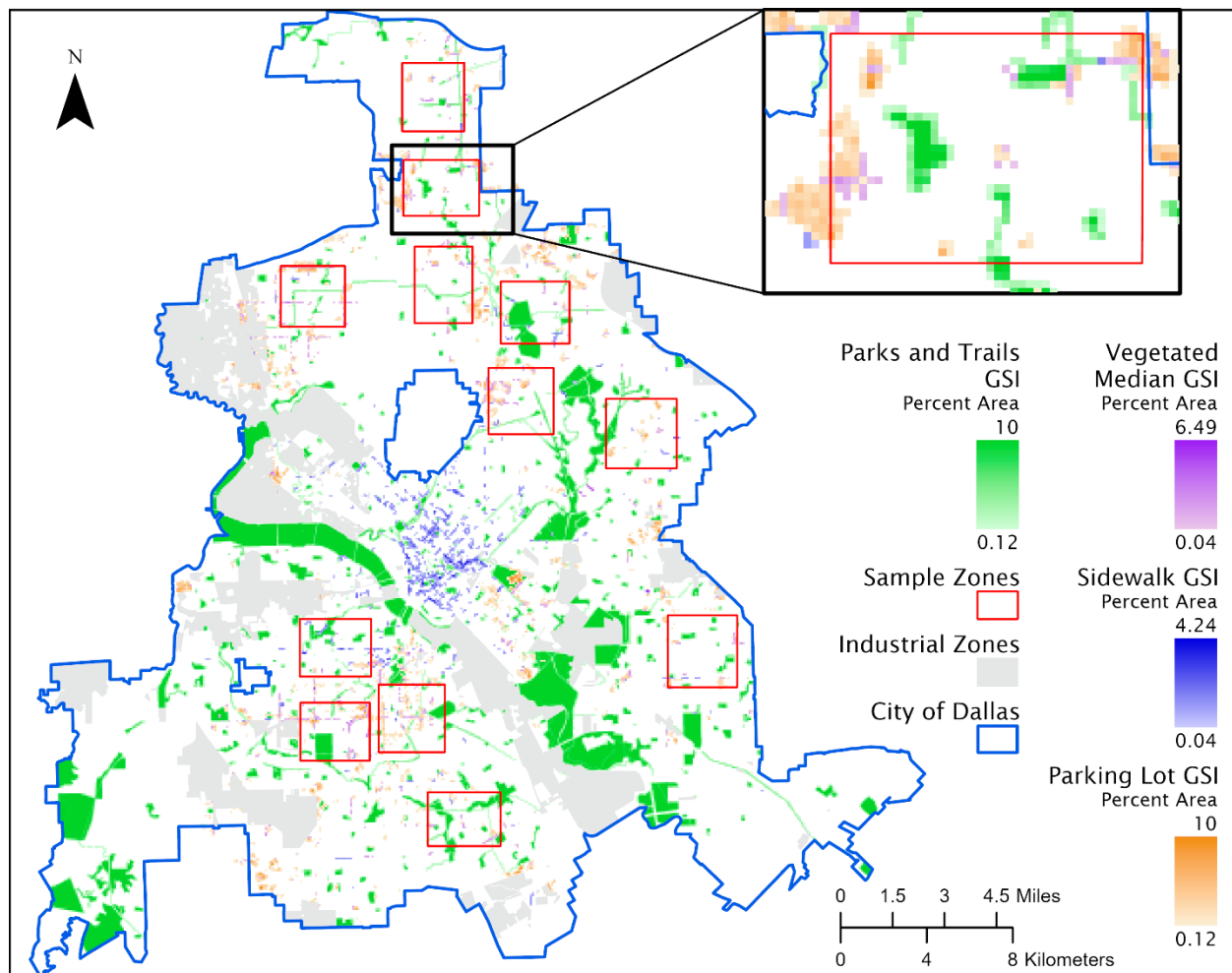


Figure 1. City of Dallas bioretention GSI candidate location type and the sample zones used for development of predictive models.

Lastly, eight sampling zones (Figure 1) were determined based on the following criteria: exclusion of major highways and interstates, outside of the intensely developed downtown area of Dallas, developed residential areas with commercial zones, inclusion of at least three of the four candidate locations. These criteria were determined due to the expected development type projected in the pilot areas as previously described. Within these zones an optimized number of sampling points per bioretention GSI type at 90 m apart were randomly generated and used to sample the data, including the percent impervious land of 2020 (Figure 2). Sampling points with no data at candidate locations were excluded, this generates 130 sidewalk samples, 300 vegetated median samples, 800 parking lot samples, and 1,000 parks and trails samples used for their respective model development (Figure 3). Finally, all remaining null values were assigned the value zero.

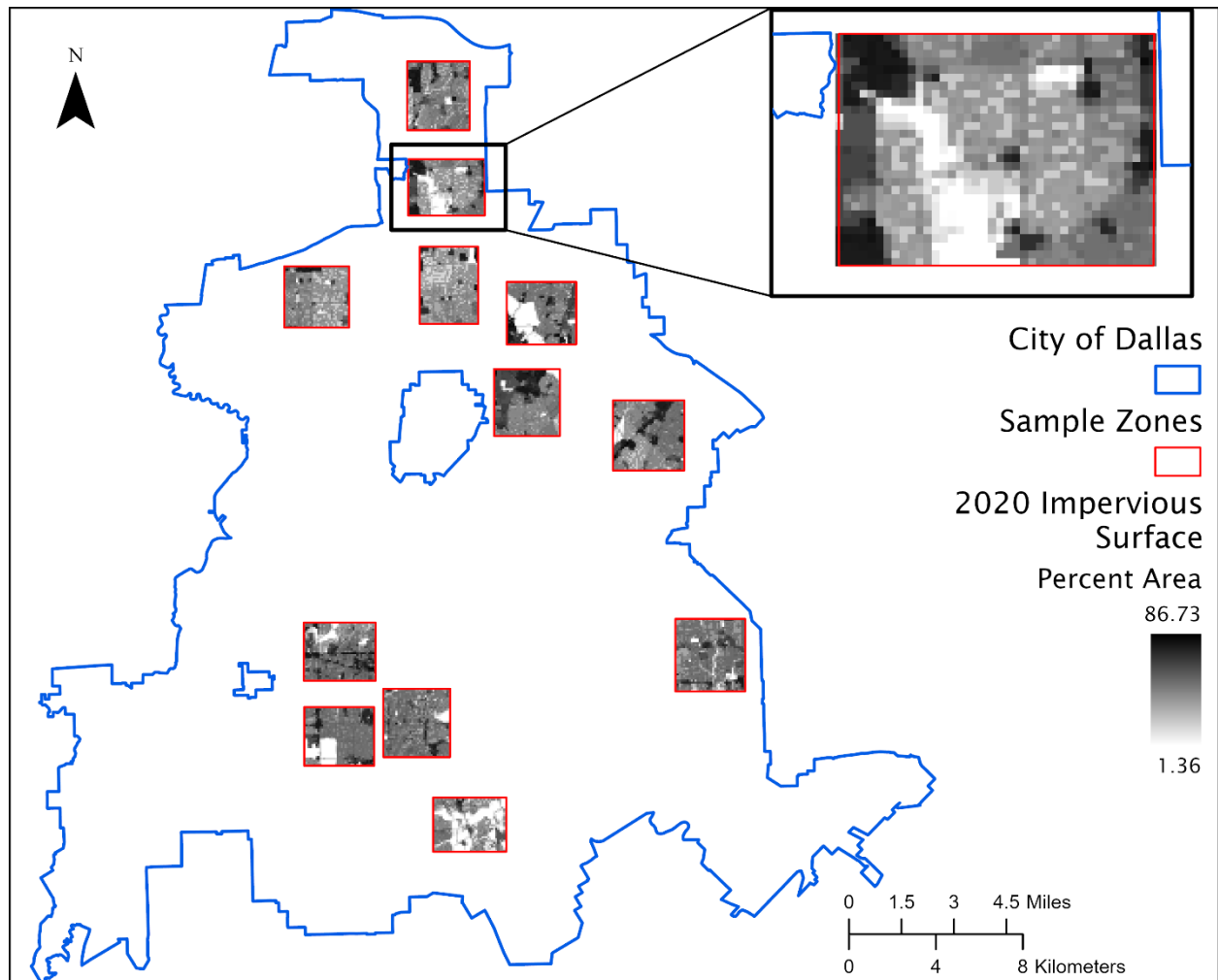


Figure 2. City of Dallas 2020 percent impervious area within the sample zones used for development of predictive models.

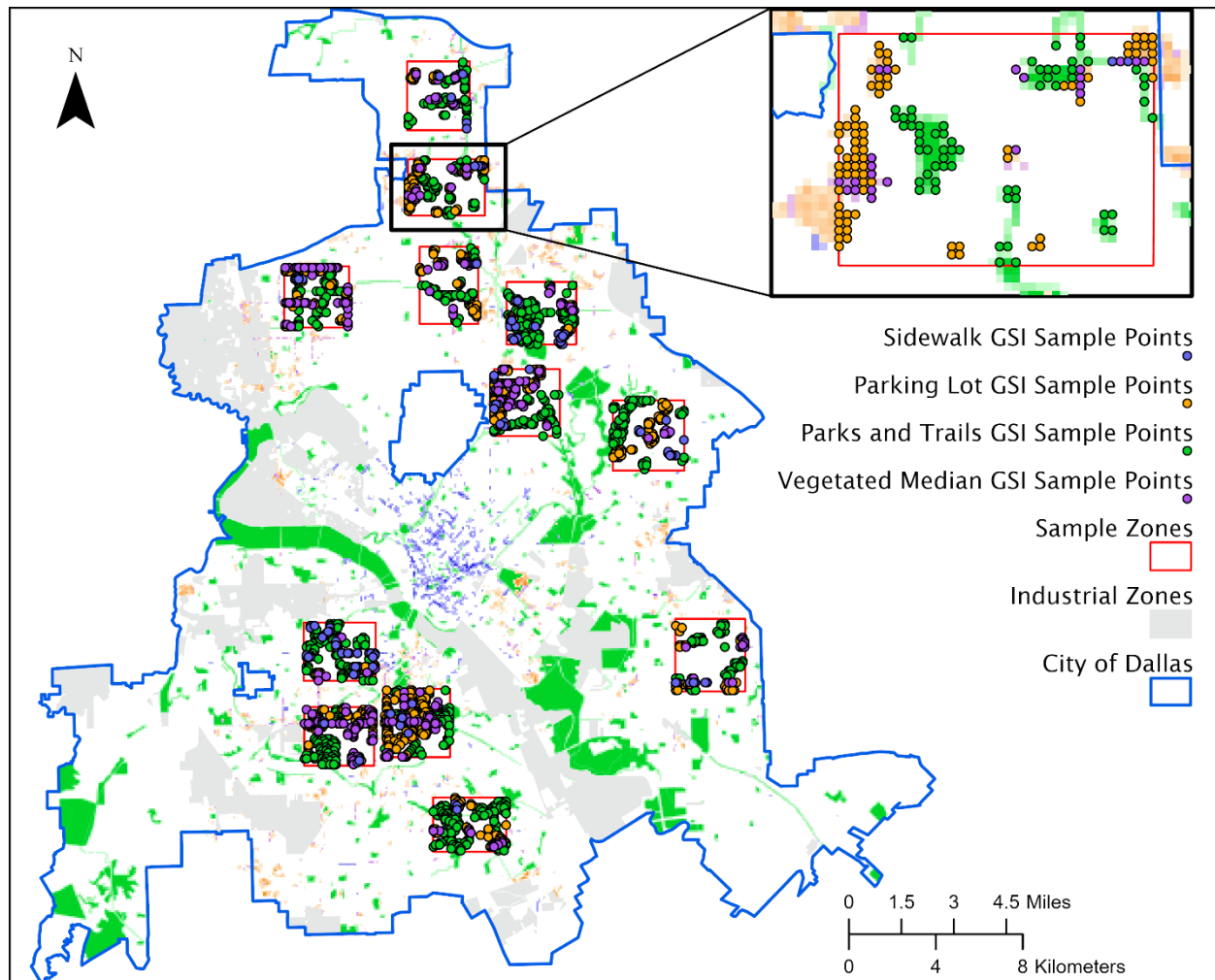


Figure 3. City of Dallas bioretention GSI candidate location type and the sampling points within the sample zones used for development of predictive models.

The linear relationship between percent impervious cover and GSI area by candidate location type are shown in Figure 4. There is very low correlation between percent impervious cover and GSI percent area cover by bioretention candidate type for vegetated medians, sidewalks and parking lots and only moderately correlated with parks and trails. These figures show high variability and relatively no consistent pattern that a general statistical model can explain. A random trees regression model is a commonly used decision-tree-based algorithm that can handle non-linear trends and include multiple explanatory variables. Random trees regression algorithm is effectively the same as a random forest regression model (RF).

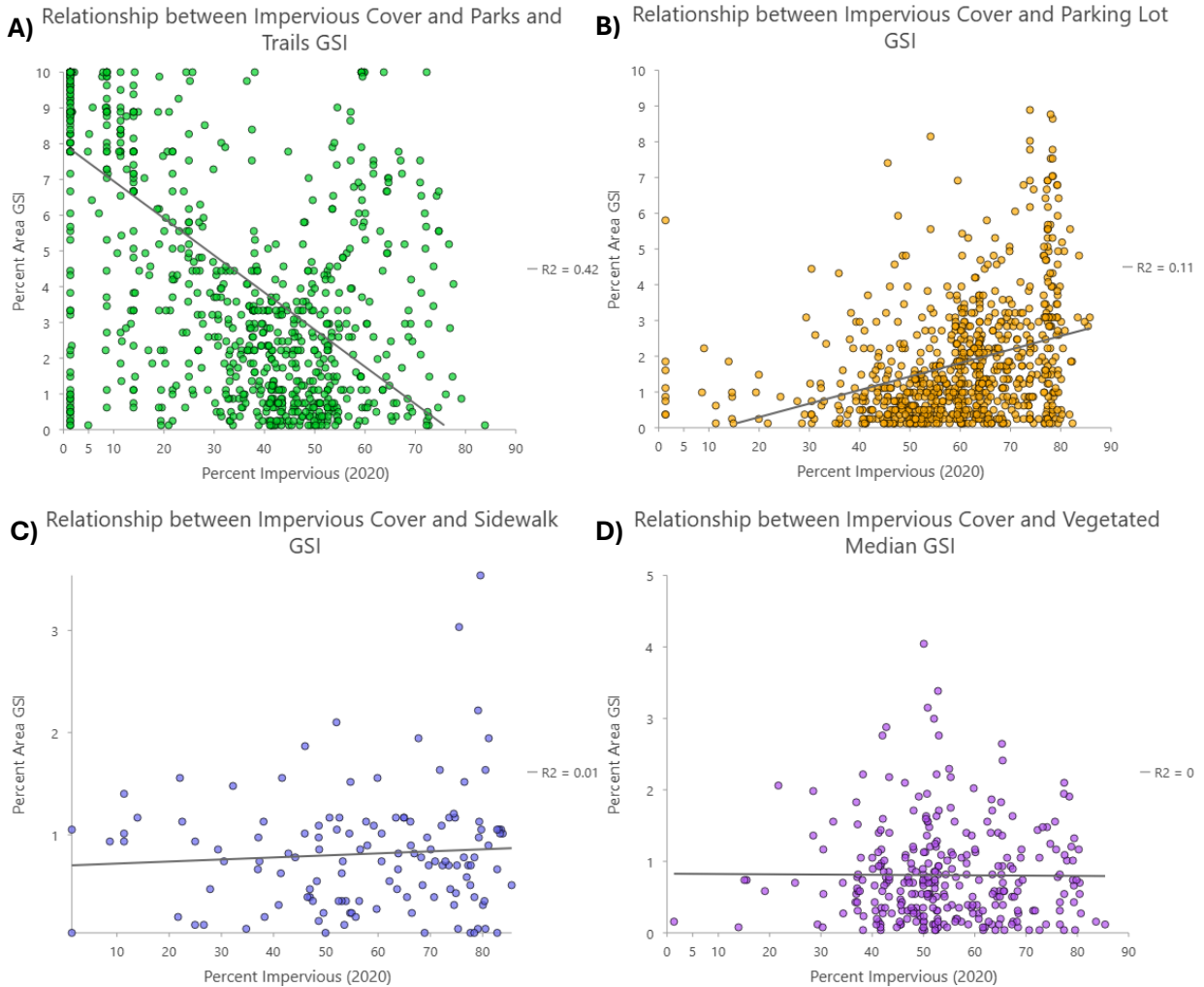


Figure 4. Linear regression between percent impervious cover (2020) and percent area of GSI by candidate location type: (A) Parks and Trails, (B) Parking Lot, (C) Vegetated Median, (D) Commercial Sidewalks at least 8 feet in width.

1.2.3 Random Forest Regression Model

Using the samples, four random trees regression models were developed to predict each candidate location type explained by percent impervious land. The RF model parameters were optimized for reduction in error, quantified as root mean square error (RMSE), this includes number of trees, number of samples, tree depth. Model training used 80% of the sample points and the other 20% were left out used for testing model predictive performance, out-of-bag (OOB) samples. Within the training data (80%), 20% of these sample points were used to test accuracy. Model performance is evaluated using coefficient of determination (R^2) that quantifies model goodness of fit (1 is a perfect fit) and the mean absolute error (MAE) that estimates the amount of uncertainty. These metrics were calculated for the training, testing, and OOB data.

1.3 Results

1.3.1 Model Performance

The model fitting performance metrics are presented in Table 1. Each model had challenges predicting the percent cover of GSI from the percent impervious cover among the test and OOB data, nearly all models had a reduction in R^2 and increased MAE compared to their training data. Thus, the variance in percent cover GSI is not strongly explained, i.e., reduced, by percent impervious cover. This is to be expected due to the high level of variance in the values observed and the number of available (1,000) sampling points, see Figure 4. For example, the model that had the best fit was for predicting the bioretention GSI percent cover in parks and trails which is in part due to the number of samples available for the model. Meanwhile, GSI in sidewalks and vegetated medians had just 130 and 300 available sample points, respectively.

Additionally, there are other explanatory variables that factor into the installation and development of GSI that are localized (e.g., soil type, localized hydrology) and others not easily measured (e.g., human behavior, policy) which will change with time and/or impact at varying spatial scales, making them difficult or impossible to use in prediction. Figure 5-7 shows the correlation between the predicted and observed (target) values by training, testing, and OOB (Test Location) data for each bioretention GSI location types. Ultimately, the models performed well enough on the training dataset for the use of these results in the context of only using percent impervious cover as the predictor of percent GSI by candidate location type. The results produced values within the range of expected values (i.e., observed), and each RF model fit the training data well.

Table 1. Model performance descriptive statistics. Coefficient of determination (R^2) and mean absolute error (MAE) values are for predicting percent impervious land 2020 used to predict the GSI percent cover of parks and trails, parking lots, vegetated medians, and sidewalks.

DESCRIPTIVE STATISTIC	PARKS AND TRAILS	PARKING LOTS	VEGETATED MEDIANS	SIDEWALKS
TRAIN MAE	1.27	0.93	0.37	0.33
TEST MAE	1.80	1.24	0.57	0.35
TRAIN R2	0.76	0.47	0.40	0.43
TEST R2	0.50	0.09	-0.07	-0.20
OOB MAE	2.13	1.04	0.62	0.43
OOB R2	0.34	0.09	-0.12	0.07

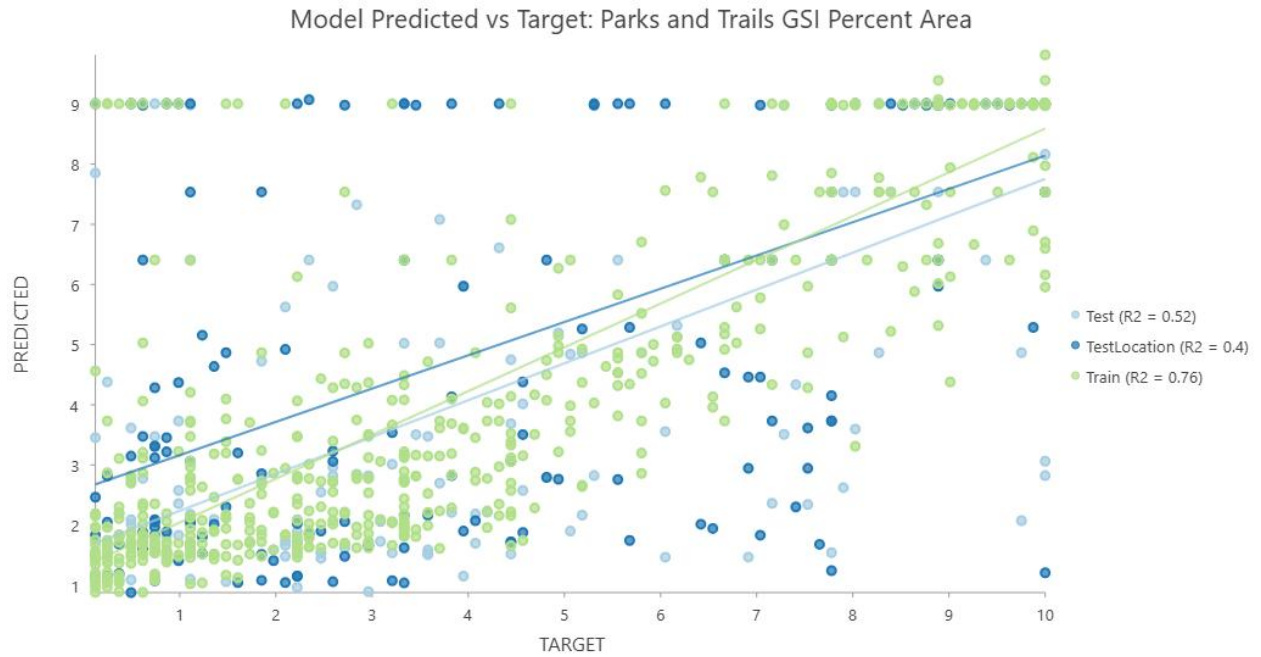


Figure 5. Parks and trails GSI percent cover linear regression between RF model predicted and target for model Testing, Test Locations, and Training data.

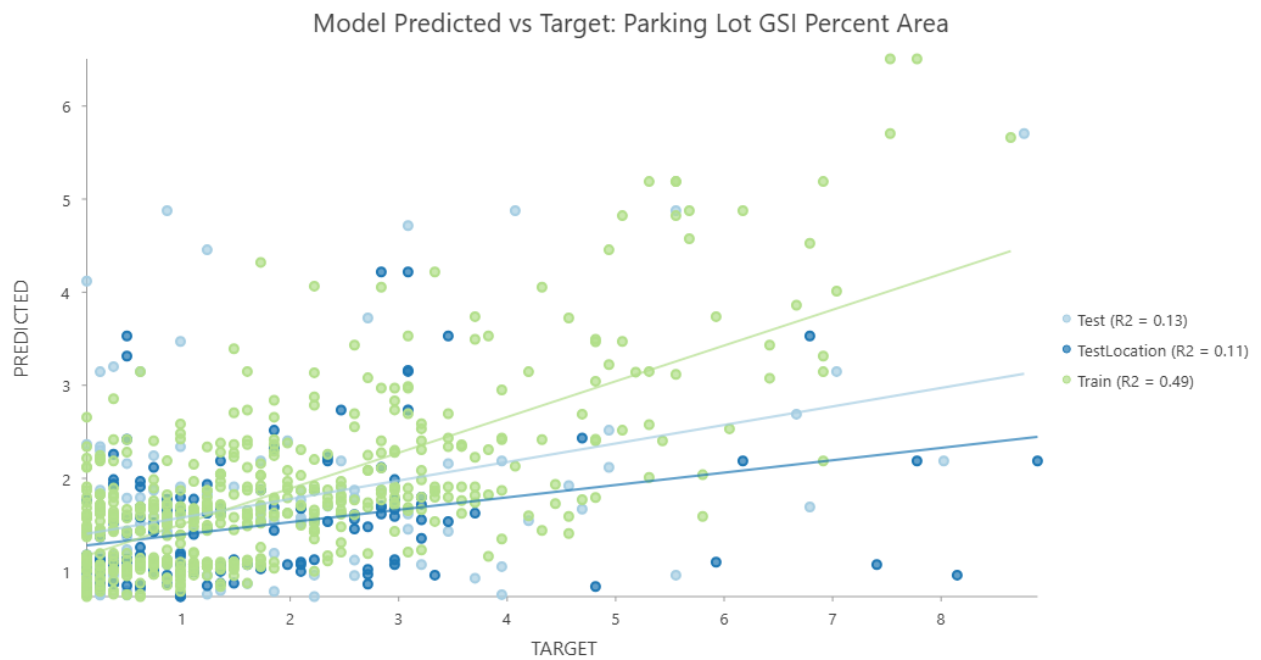


Figure 6. Parking lot GSI percent cover linear regression between RF model predicted and target for model Testing, Test Locations, and Training data.

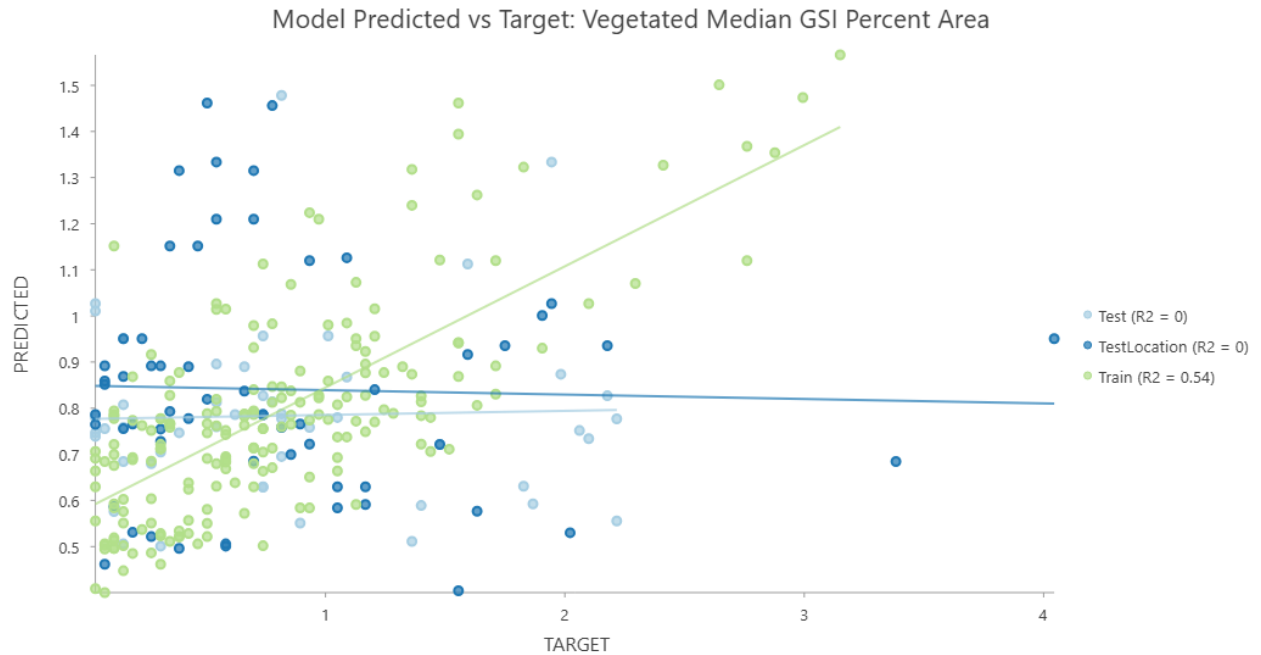


Figure 7. Vegetated median GSI percent cover linear regression between RF model predicted and target for model Testing, Test Locations, and Training data.

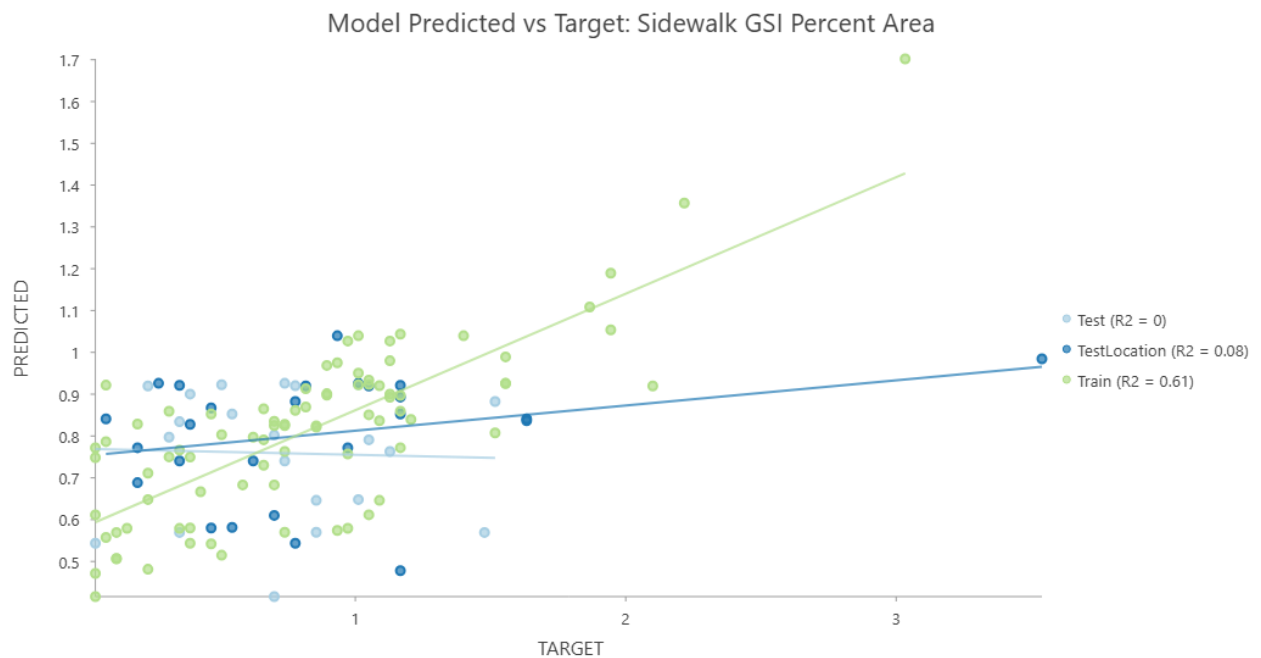


Figure 8. Sidewalk GSI percent cover linear regression between RF model predicted and target for model Testing, Test Locations, and Training data.

1.3.2 Estimated Maximum Bioretention

The maximum potential area of bioretention by candidate type was estimated by the derived models for 2045 and 2070. Due to the limitations of the analysis, the area of

bioretention in parks and trails was inflated by the correlation of rural area crops and rangeland with low impervious cover. To reduce the influence of rural area lands in the results, the area of the watershed was isolated to the urban-suburban areas, defined as at least 10% impervious cover. This also reduced the lake bias on the results, which were important when calculating the maximum potential area of bioretention GSI in parks and trails. The use of 10% impervious cover was used based on a review by Brabec, et al. (2002) identified an impact threshold for biotic measurements of 3.6-15% impervious surface, rather than the 20% impervious surface used in MRLC LCLU classification. Additionally, when incorporating GSI into hydrological models there is a limitation of applying these to urban land uses.

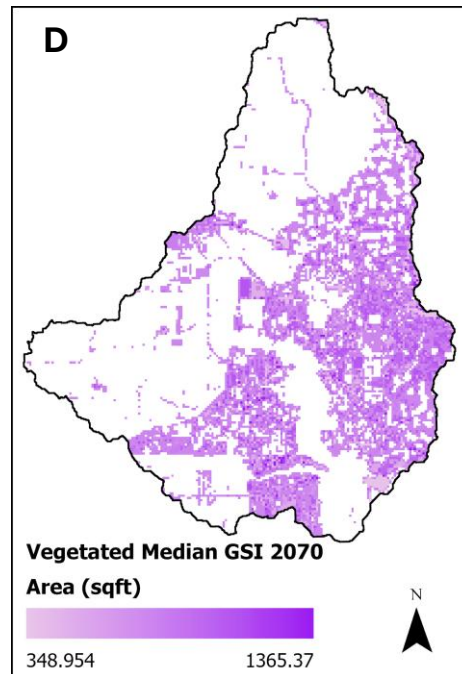
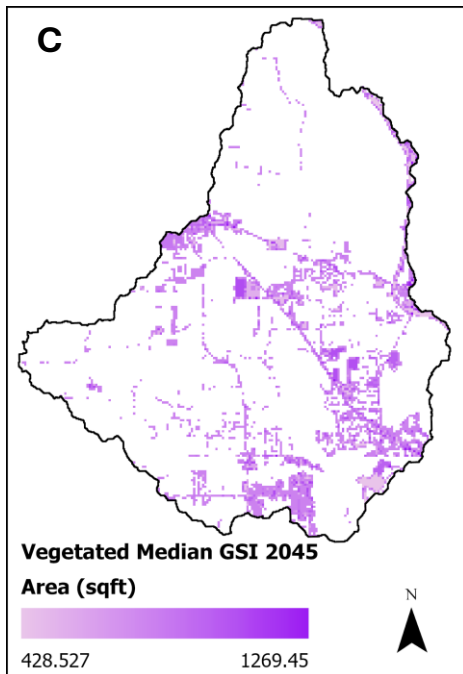
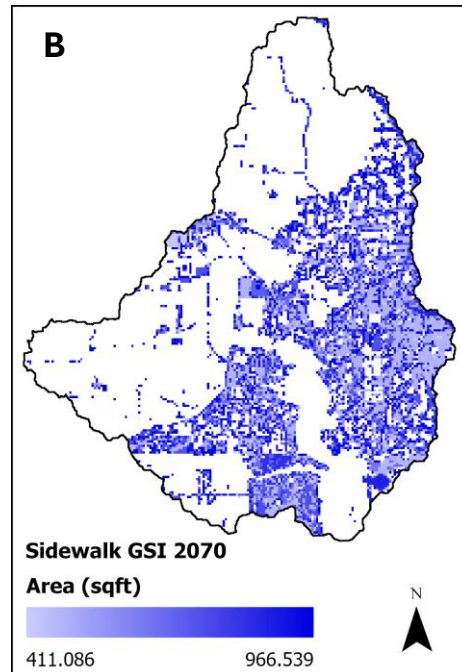
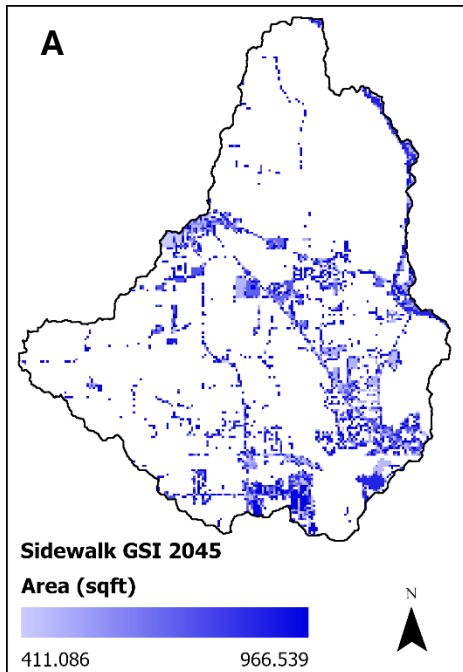


Figure 9. Estimated maximum potential area (ft^2) of bioretention by location type in the Eagle Mountain pilot watershed for 2045 and 2070. For all maps the area increases from light to dark along the color gradient. A and B show the GSI area of sidewalks at least 8ft in width; C and D show the GSI area of vegetated medians; E and F show the GSI area of parking lots; and G and H show the GSI area of parks and trails.

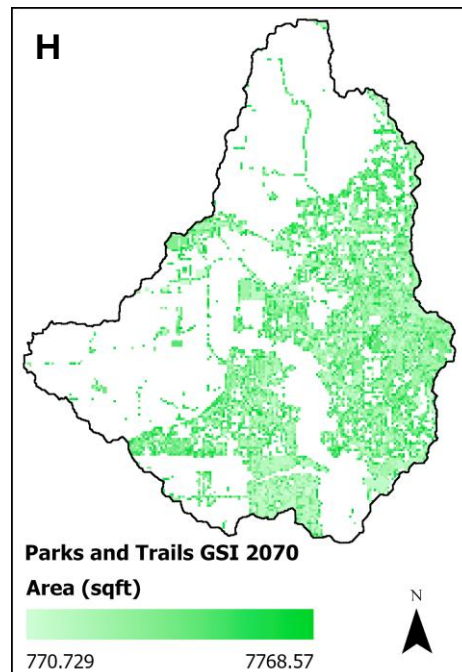
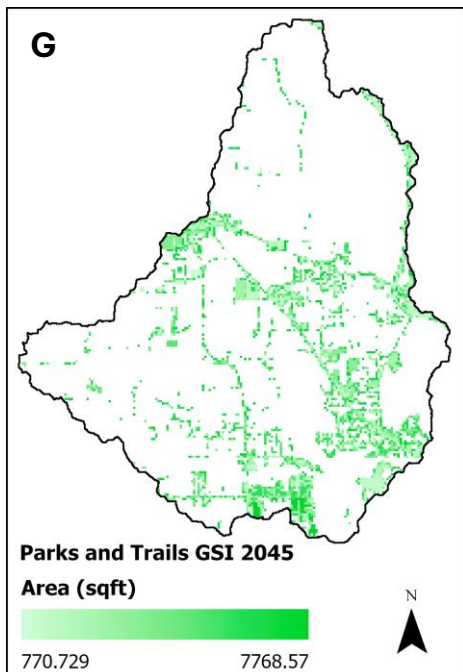
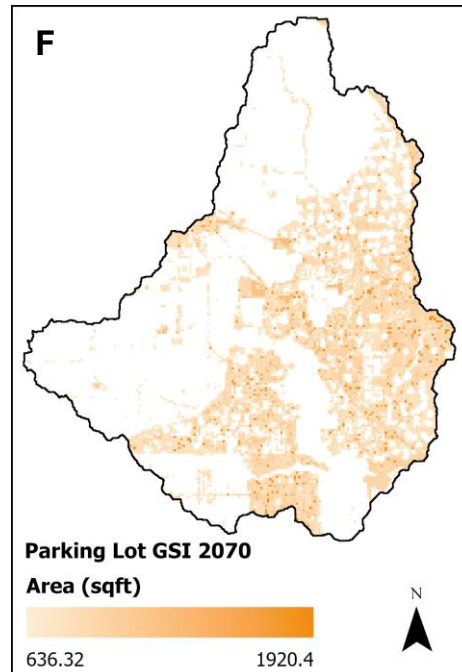
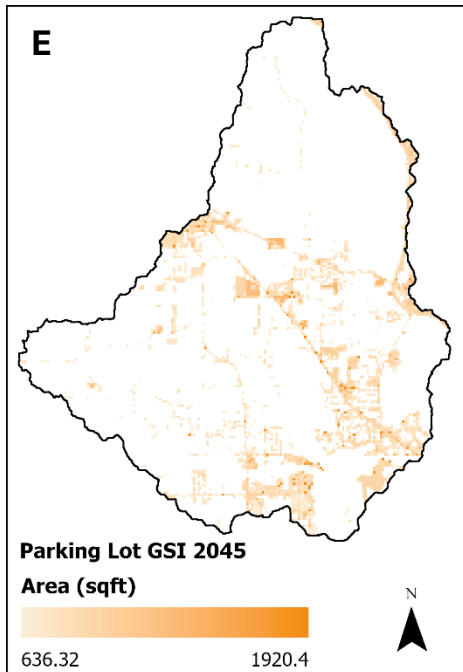


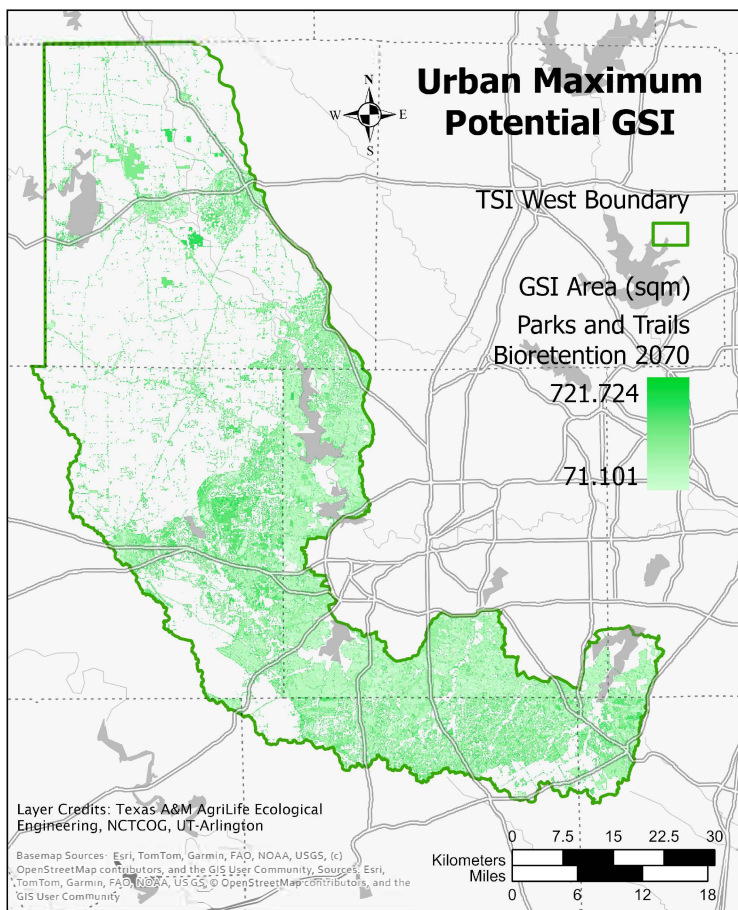
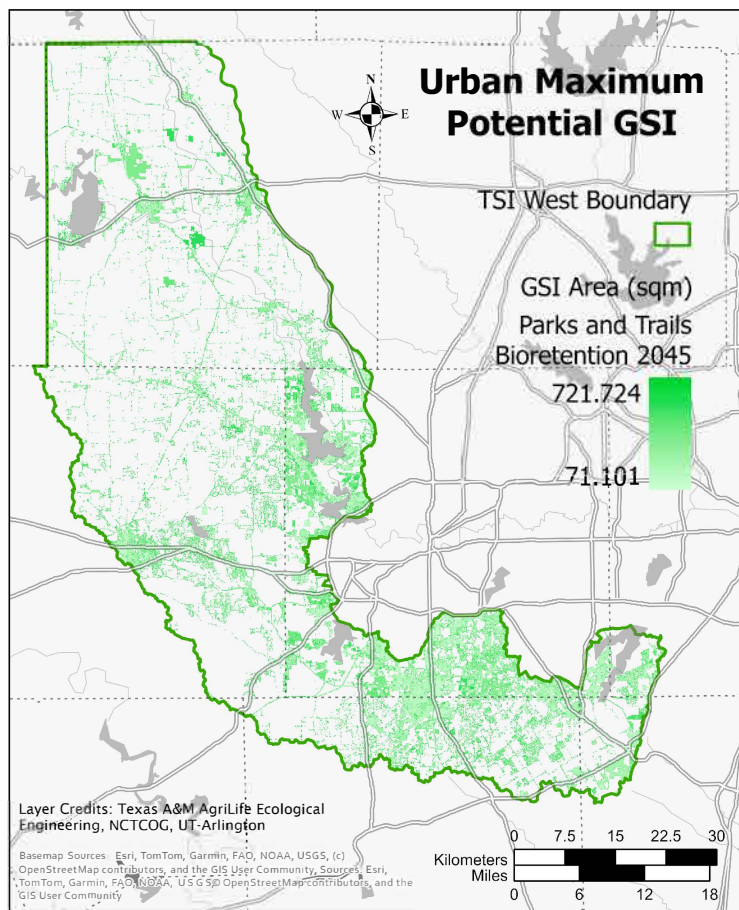
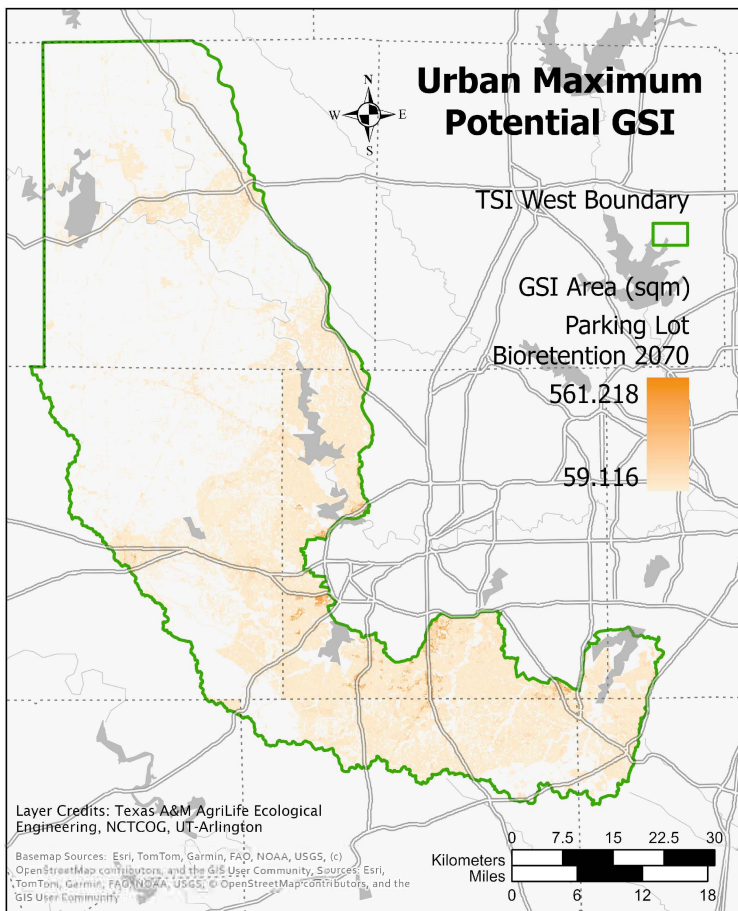
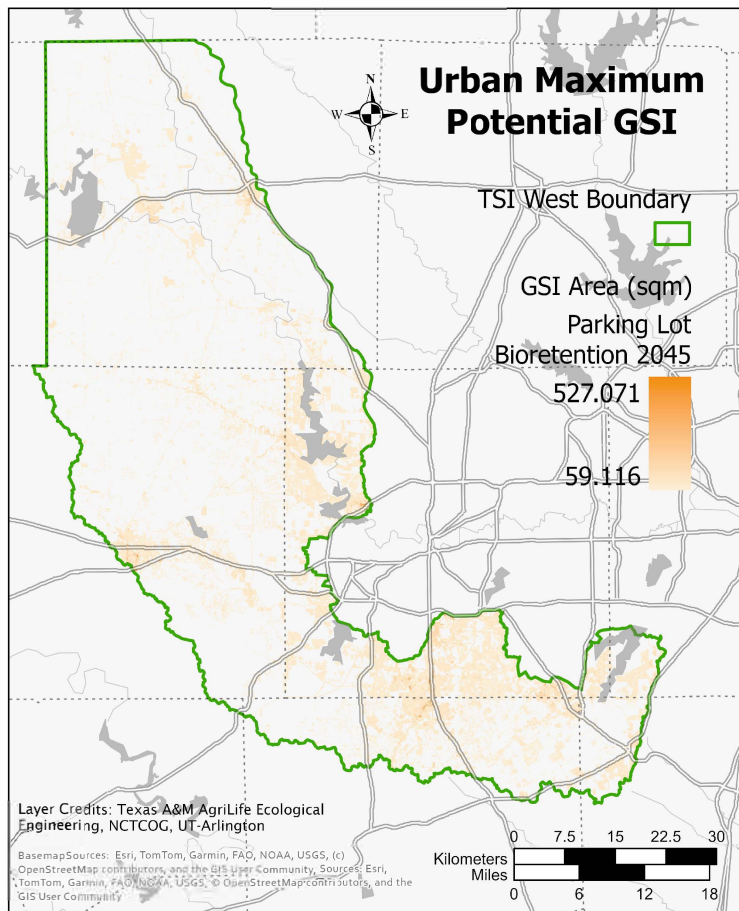
Figure 10. Cont. Estimated maximum potential area (ft^2) of bioretention by location type in the Eagle Mountain pilot watershed for 2045 and 2070. For all maps the area increases from light to dark along the color gradient. A and B show the GSI area of sidewalks at least 8ft in width; C and D show the GSI area of vegetated medians; E and F show the GSI area of parking lots; and G and H show the GSI area of parks and trails.

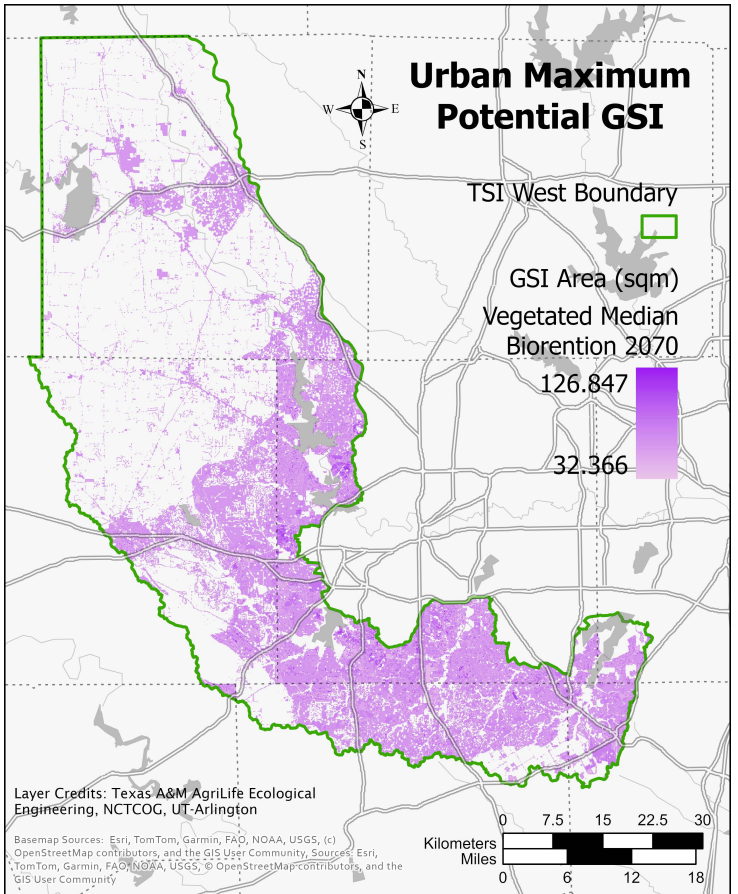
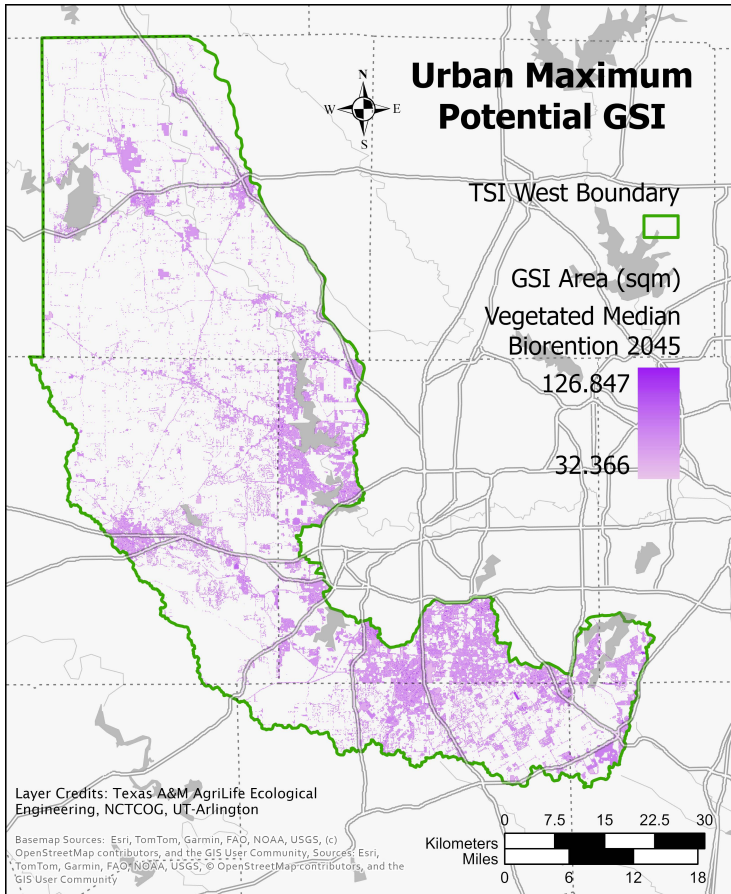
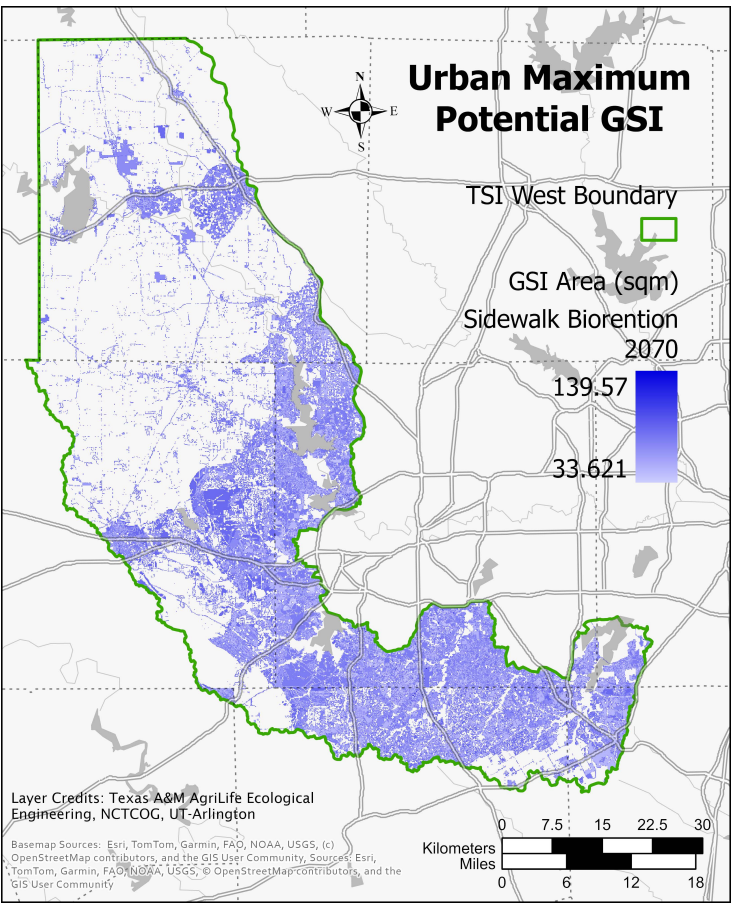
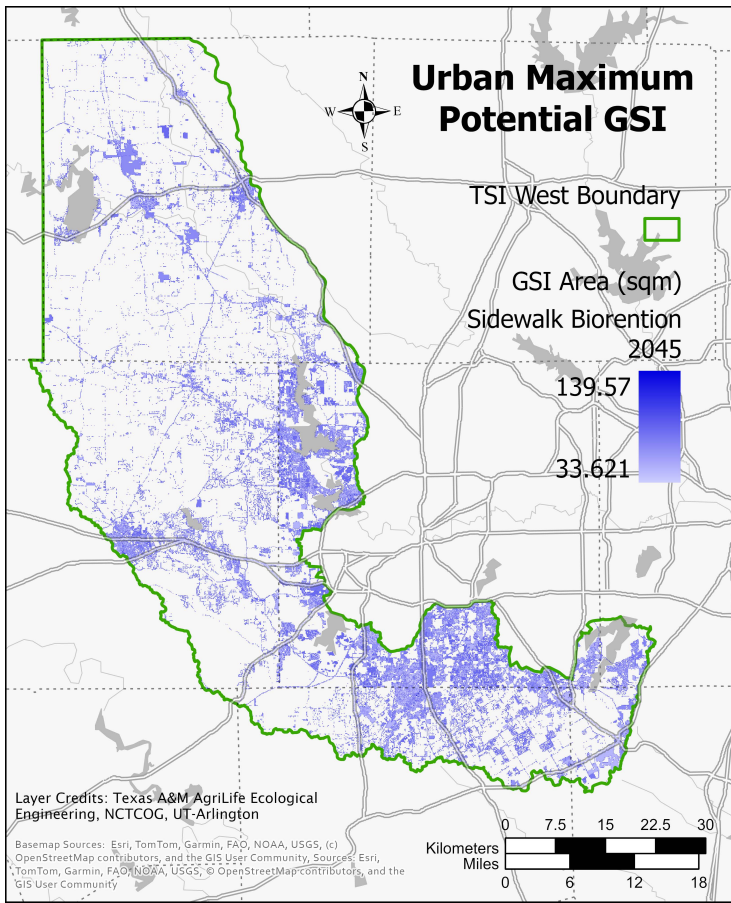
The summary of these estimated areas for the Eagle Mountain pilot watershed is shown in Table 2. In the urban areas of the watershed total area of maximum bioretention GSI

increased from 2045 to 2070. These predictions suggest that for 2045 approximately 27.96% of the urban area in the watershed could hold bioretention, and approximately 23.46% for 2070.

Table 2. Summary of predicted maximum potential bioretention area (ft²) within the Eagle Mountain pilot watershed area by candidate location type. The GSI area is calculated within urban areas defined as at least 10% impervious land cover.

Candidate Locations	Bioretention criteria	2045 Maximum bioretention area (ft²)	2070 Maximum bioretention area (ft²)
Parks and Trails	10%	14,117,487.66	30,636,846.88
Parking Lots	10%	3,210,058.69	8,224,577.66
Vegetated Medians	35%	2,494,219.56	6,252,554.50
Sidewalks	35%	2,828,478.53	6,446,865.16
Total Bioretention GSI		22,650,244.44	51,560,844.20
Total Urban Area in Watershed		81,021,819.57	219,771,740.51
Total Watershed Area		2,152,404,040.70	2,152,404,040.70





TSI West study area predicted bioretention GSI candidate locations area in square meters.

YEAR	SIDEWALKS	PARKING LOTS	PARKS AND TRAILS	VEGETATED MEDIANS	TOTAL
2045	7,595,958	9,178,993	36,566,025	6,821,199	60,162,175
2070	13,275,617	17,983,810	57,343,548	12,610,014	101,212,989