Data Shop

Data Shop, a department of Cityscape, *presents short articles or notes on the uses of data in housing and urban research. Through this department, the Office of Policy Development and Research introduces readers to new and overlooked data sources and to improved techniques in using well-known data. The emphasis is on sources and methods that analysts can use in their own work. Researchers often run into knotty data problems involving data interpretation or manipulation that must be solved before a project can proceed, but they seldom get to focus in detail on the solutions to such problems. If you have an idea for an applied, data-centric note of no more than 3,000 words, please send a one-paragraph abstract to datashop@hud.gov for consideration.*

Mapping Manufactured Housing Nationwide

Noah J. Durst
Nithya Mylakumar
Angela Perez
Michigan State University

Angelica Gacis
Data Science Consultant

Abstract

This study uses building footprints from Microsoft and OpenStreetMap and the Python package momepy *to measure the shape, size, and placement of buildings and their 5, 10, and 20 nearest neighbors across the continental United States. Using estimates of building and neighborhood morphology and machine learning, we predict whether each building is a singlewide manufactured home and whether it is in a manufactured home park, informal or manufactured home subdivision, or another setting. We describe the methods used to create these predictions and discuss issues of model performance and their implications for future research, compare our estimates with the locations of manufactured homes documented in the American Community Survey and with government and private registries of these communities, illustrate their distribution nationwide, and present descriptive statistics on their demographic and socioeconomic characteristics. Our findings illustrate that manufactured home parks are more common in Midwestern and Northeastern states, whereas informal or manufactured home subdivisions are more common in Southern and Western states. We find that both neighborhoods are demographically diverse but economically disadvantaged. We conclude by briefly discussing the implications of our research for state and federal housing policy.*

Introduction

Manufactured homes are one of the most affordable housing options produced today and are home to approximately 1.7 million renter households and 5.3 million owner households (Durst and Sullivan, 2019). Approximately one-third of these manufactured homes are in parks that are owned and operated by landlords (Durst and Sullivan, 2019; Sullivan, 2017), although cooperative, resident-owned models also exist (Mukhija and Mason, 2014). A substantial share of manufactured homes is also in residential subdivisions, where residents often own both the housing unit and the land on which it sits (Durst and Sullivan, 2019). Such neighborhoods come in two broad types—manufactured home subdivisions, where manufactured housing makes up most of the housing (Durst et al., 2021), and informal subdivisions, where manufactured housing makes up a substantial minority of the dwellings (Reyes et al., 2024). Manufactured home subdivisions have received little attention in academic literature. On the other hand, informal subdivisions have been widely studied in the U.S.-Mexico border region (Ward, 1999). More recently, scholars have used a variety of methods to identify and examine the prevalence of informal subdivisions across the country, although these analyses are all restricted to a small sample of locations or can only capture variation in the prevalence of these neighborhoods at the regional level (Durst and Sullivan, 2019; Reyes et al., 2024; Ward and Peters, 2007).

Although residents in manufactured housing often have lower rates of household mobility than traditional renters (Boehm and Schlottmann, 2006), they face distinct threats to tenure security depending on their land tenure status. In parks, threats often involve the risk of park closure and redevelopment (Sullivan, 2017). In subdivisions, threats include exploitative lending practices, unclear or insecure property titles, and high rates of foreclosure or repossession (Olmedo and Ward, 2016; Ward, 1999). Residents in both communities also face elevated environmental hazards because of their disproportionate locations in high-risk areas, such as flood plains, and the poor quality of infrastructure (Durst, 2016; Pierce, Gabbe, and Gonzalez, 2018; Reyes et al., 2024; Rumbach, Sullivan, and Makarewicz, 2020).

The federal and multiple state governments maintain registries to track housing and infrastructure conditions in these neighborhoods and ensure rapid response in the case of storms or other natural hazards. The U.S. Department of Homeland Security (DHS; 2023) Homeland Infrastructure Foundation-Level Database documents the longitude and latitude of more than 56,000 manufactured home communities.[1] The database refers to these communities as *parks*, although Durst et al. (2025) document that a substantial fraction in Texas (15 percent) are, in fact, manufactured home subdivisions. The federal government also monitors housing and infrastructure conditions in informal subdivisions developed before 1990 and within 150 miles of the U.S.-Mexico border. These neighborhoods are referred to as *colonias* under state and federal law (Ward, 1999). The U.S. Department of Housing and Urban Development (HUD) database tracks the longitude and latitude of colonias across the border region,[2] and the Texas Office of the Attorney General's (2023) colonias database documents the neighborhood boundaries for pre-1990 colonias

[1] https://hifld-geoplatform.hub.arcgis.com/.

[2] https://www.hudexchange.info/programs/cdbg-colonias/.

within the Texas border region.[3] However, to our knowledge, no active government database tracks the location of informal subdivisions developed after 1989 or outside the 150-mile border region. In this study, we use building footprints and machine learning to conduct the first national, subregional analysis of the prevalence and characteristics of these communities.

Data and Methods

This study uses building footprint data to document the locations and outlines of buildings across the continental United States. We use these building footprints to measure the size, shape, and placement of each building, as well as their 5, 10, and 20 nearest neighbors. We then use a sample of approximately 45,000 buildings in Harris County, Texas, to develop machine learning models capable of predicting whether a building footprint is a singlewide manufactured housing unit. Then, using a sample of more than 500,000 building footprints across 19 Texas counties, we develop models to predict whether a building is within a manufactured home park or an informal or manufactured home subdivision. To examine model performance, we use a test dataset containing validated singlewide manufactured homes and buildings in validated parks and subdivisions to calculate accuracy, precision, and recall. We also examine out-of-sample model performance by comparing our predictions with prior research, data on the location of manufactured housing units from the American Community Survey, and existing park and subdivision registries.

We use two publicly accessible datasets of building footprints from Microsoft (2022) and OpenStreetMap (OSM; 2023) to identify building footprints nationwide. The accuracy and coverage of building footprint data vary considerably by both source and region (Gonzalez, 2023). Where a particular building footprint is available in both datasets, we use the footprints delineated by OSM, which are often manually reviewed prior to publication or derived from validated data sources (Biljecki, Chow, and Lee, 2023). We document the location of manufactured home parks using validated locations compiled by Durst et al. (2025) and Sullivan, Makarewicz, and Rumbach (2022). We document the location of informal or manufactured home subdivisions using parcel records from the Texas Natural Resources Information System Datahub.[4]

We use the R and Python languages for data collection, cleaning, and analysis and the open-source QGIS software for data exploration and visual verification. We begin by calculating selected morphometrics for each building (the rows labeled "Self" in the "Universe" column in exhibit 1), including the area, perimeter, length-width ratio, and distance to the nearest building for all building footprints in the continental United States, using the *momepy* package in Python (Fleischmann, 2019). To capture the morphology of the local neighborhood, we also measure the median and interquartile ranges among the nearest 5, 10, and 20 buildings surrounding each individual building footprint in the nation for each of these morphometrics. We use these morphometrics to conduct two different supervised classification efforts. First, we predict whether a building is a singlewide manufactured home. We focus specifically on singlewide manufactured homes because their distinct size and shape make them easier to identify via analysis of building footprints. Second, we predict whether buildings are in a manufactured home park or an informal or manufactured home

[3] https://www.texasattorneygeneral.gov/colonias.

[4] https://data.geographic.texas.gov.

subdivision. Although we briefly describe the prediction of singlewide manufactured homes, the focus of this article is primarily on the prediction of buildings in parks and subdivisions.

Exhibit 1

Building and Neighborhood Morphometrics

Variable Name	Measurement	Summary Statistic	Universe	Manufactured Home Parks	Informal or Manufactured Home Subdivisions	Other Locations
	Building Area		Self	117.58	162.57	322.90
area_m_5	Building Area	Median	5 Neighbors	112.55	153.62	261.45
area_m_10	Building Area	Median	10 Neighbors	112.92	152.84	253.55
area_m_20	Building Area	Median	20 Neighbors	114.11	153.07	246.98
area_r_5	Building Area	IQR	5 Neighbors	27.66	65.58	97.52
area_r_10	Building Area	IQR	10 Neighbors	40.41	94.24	130.48
area_r_20	Building Area	IQR	20 Neighbors	46.18	101.29	130.16
	Perimeter		Self	50.79	54.43	70.47
peri_m_5	Perimeter	Median	5 Neighbors	50.90	53.72	66.55
peri_m_10	Perimeter	Median	10 Neighbors	51.26	53.89	66.09
peri_m_20	Perimeter	Median	20 Neighbors	51.62	53.99	65.57
peri_r_5	Perimeter	IQR	5 Neighbors	6.86	12.34	12.85
peri_r_10	Perimeter	IQR	10 Neighbors	9.87	17.41	17.75
peri_r_20	Perimeter	IQR	20 Neighbors	10.92	18.48	18.39
	Distance		Self	6.82	16.63	9.12
nn_d_m_5	Distance	Median	5 Neighbors	6.12	13.61	6.89
nn_d_m_10	Distance	Median	10 Neighbors	6.03	13.16	6.48
nn_d_m_20	Distance	Median	20 Neighbors	6.03	12.90	6.23
nn_d_r_5	Distance	IQR	5 Neighbors	2.07	8.42	3.88
nn_d_r_10	Distance	IQR	10 Neighbors	2.94	12.43	5.43
nn_d_r_20	Distance	IQR	20 Neighbors	3.31	13.61	5.66
	Length/Width		Self	3.76	2.73	2.43
LW_r_m_5	Length/Width	Median	5 Neighbors	3.78	2.56	2.34
LW_r_m_10	Length/Width	Median	10 Neighbors	3.78	2.51	2.32
LW_r_m_20	Length/Width	Median	20 Neighbors	3.74	2.48	2.31
LW_r_r_5	Length/Width	IQR	5 Neighbors	0.70	0.58	0.28
LW_r_r_10	Length/Width	IQR	10 Neighbors	1.06	0.83	0.40
LW_r_r_20	Length/Width	IQR	20 Neighbors	1.19	0.85	0.42
	Singlewide		Self	0.56	0.10	0.00
MH_foot_5	Singlewide	Percentage	5 Neighbors	0.56	0.11	0.00
MH_foot_10	Singlewide	Percentage	10 Neighbors	0.55	0.11	0.01
MH_foot_20	Singlewide	Percentage	20 Neighbors	0.53	0.10	0.01

Notes: This table presents morphometrics for buildings in manufactured home parks, informal or manufactured home subdivisions, and other locations. We calculate the building footprint area (square meters), length of the building footprint perimeter (meters), distance to the nearest building (meters), and the ratio of the length of the building to the width of the building. We also calculate the median and interquartile range for each of these building characteristics among the 5, 10, and 20 nearest neighbors. We use these building characteristics to predict whether a building is a singlewide manufactured home. We then calculate the percentage of buildings among the 5, 10, and 20 nearest neighbors that are singlewide homes.

Sources: These data were produced using building footprints acquired from Microsoft (2022) and OpenStreetMap (2023), parcel data from the Texas Natural Resources Information System Datahub, and the locations of manufactured home parks compiled by Durst et al. (2025) and Sullivan, Makarewicz, and Rumbach (2022)

Predicting Singlewide Manufactured Homes

We begin by training a classifier to identify singlewide manufactured housing units. Singlewide manufactured homes are constructed on a single chassis and designed to be transported to a site in one piece. They are smaller than doublewide units, which are constructed on two separate chassis and transported independently to the site. Singlewide homes typically have widths of approximately 5 to 6 meters and lengths up to approximately 25 meters, although the sizes of the building footprint polygons as delineated in the Microsoft and OSM datasets often vary considerably because of building modifications (for example, awnings and carports) or errors in the delineation of the building footprints. To develop a dataset of validated singlewide manufactured homes, we use a sample of more than approximately 45,000 buildings in Harris County (Houston), Texas. To validate the locations of singlewide manufactured homes, we use QGIS software to review satellite imagery, building footprints, parcel records, and land use classifications indicative of manufactured housing from the Harris County Appraisal District. Two researchers have reviewed each building footprint within the sample. Buildings are classified as singlewide manufactured homes if on a parcel with a land use code indicative of manufactured housing or within a manufactured home park identified by Sullivan, Makarewicz, and Rumbach (2022) and had dimensions that approximated those described previously.[5] We then use the building area, perimeter length, length-width ratio, and distance to the nearest neighbors and the median and interquartile range of each of these morphometrics for the 5, 10, and 20 nearest neighbors and a Random Forest classifier to predict whether each building is a singlewide manufactured home. Within the test dataset, our models achieve an accuracy, precision, and recall of more than 0.98, 0.92, and 0.86, respectively. At the county level nationwide, the percentage of buildings predicted to be singlewide manufactured homes is highly correlated with the percentage of housing units that are manufactured homes (correlation coefficient of 0.5). As the next section describes, we use these predictions of singlewide manufactured homes as independent variables in models predicting whether the building is in a park, subdivision, or other location.

Predicting Parks and Subdivisions

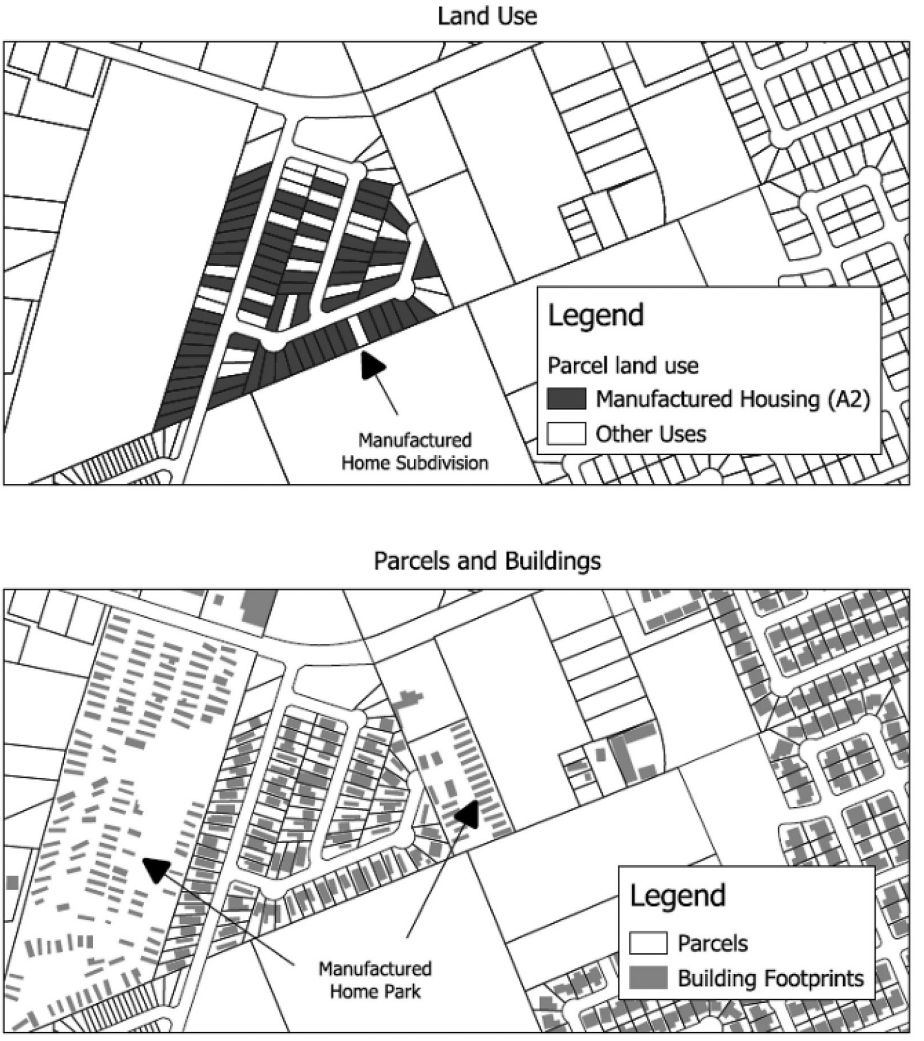
We now turn to a discussion of our primary objective in this study—predicting whether a building is in a manufactured home park or an informal or manufactured home subdivision. We begin by overlaying parcel records and building footprints, as exhibit 2 illustrates. We identify the location of each building by joining parcel polygons to building centroids. To identify manufactured home parks, we use a dataset compiled by Durst et al. (2025) of approximately 1,500 validated parcels containing manufactured home parks across 19 Texas counties. We treat any building on a validated park parcel compiled by Durst et al. (2025) as a manufactured home park. To identify informal and manufactured home subdivisions, we use county property records and parcel boundaries to identify parcels with a land use classification indicative of manufactured housing on residential land (A2). We treat any building as an informal or manufactured home subdivision for

[5] Because of building additions or errors in the delineation of building footprints, not all building footprints for our validated singlewide manufactured homes are rectangular. However, assuming a rectangular footprint, the area, length, and width have the following relationship: area = length * width. In our training dataset, the average singlewide home has an area of 111 square meters (1,200 square feet) and a length-width ratio of approximately 4. For rectangular buildings, it equates to an average length of 21 meters (69 feet) and an average width of 5 meters (17 feet).

which 1 or more of the 10 nearest buildings is on a parcel with an A2 land use classification. To eliminate any large parcels next to but not within residential subdivisions, we remove buildings on parcels that are twice as large as the average of the nearest 10 parcels.

Exhibit 2

Parcel and Building Datasets



Sources: These maps were produced using building footprints acquired from Microsoft (2022) and OpenStreetMap (2023) and parcel data from the Texas Natural Resources Information System Datahub

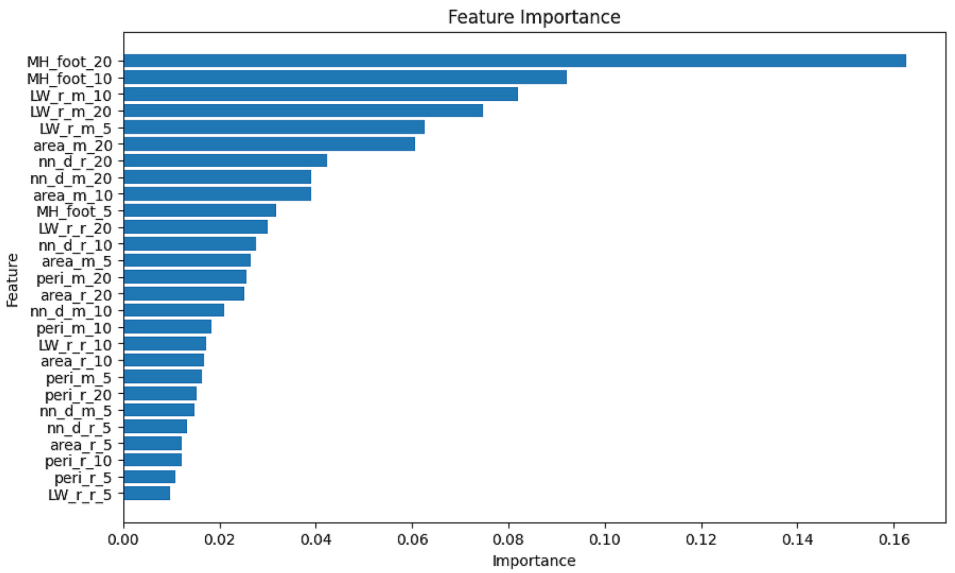
In exhibit 1, we present the average for each morphometric for manufactured home parks, informal or manufactured home subdivisions, and other locations. To reiterate, we calculate the building area, perimeter length, length-width ratio, distance to nearest neighbors, and the median and interquartile ranges of these morphometrics for the 5, 10, and 20 nearest neighbors. We also use the results from our first stage predictions to calculate the percentage of buildings that are predicted to be singlewide

manufactured homes. These building characteristics illustrate that buildings in manufactured home parks are smaller (117 square meters) than buildings in informal or manufactured home subdivisions (162 square meters) or other locations (322 square meters), are closer to neighboring buildings (6.8 versus 16.6 and 9.1 meters, respectively), and have a higher length-width ratio (3.76 versus 2.73 and 2.43, respectively). Similarly, singlewide homes constitute 56 percent of buildings in manufactured home parks, 10 percent of buildings in informal or manufactured home subdivisions, and less than 1 percent of buildings in other locations.

We then apply machine learning models to predict the type of community in which each building is located. Our training and test dataset is drawn from a sample of more than 5 million buildings across 19 Texas counties analyzed by Durst et al. (2025). Given the size of the dataset, we select a 10-percent sample of approximately 550,000 buildings for training and testing purposes. We divide this sample into a training set (80 percent) and a test set (20 percent). To address class imbalance, we apply the Synthetic Minority Oversampling Technique. We train a Random Forest classifier and validate the model using five-fold cross-validation on the training set. Specifically, we divide the training set into five folds, with four folds used for training and the fifth used for validation in each iteration. We repeat cross-validation five times, after which we evaluate the model on the independent test set. Exhibit 3 displays the feature importance of the final model. The share of singlewide homes among the 10 or 20 nearest neighbors and the median length-width ratio among the 5, 10, or 20 nearest neighbors are the five most important morphometrics. The unique, elongated shape of singlewide manufactured homes is central to distinguishing parks and subdivisions from buildings in other locations.

Exhibit 3

Feature Importance

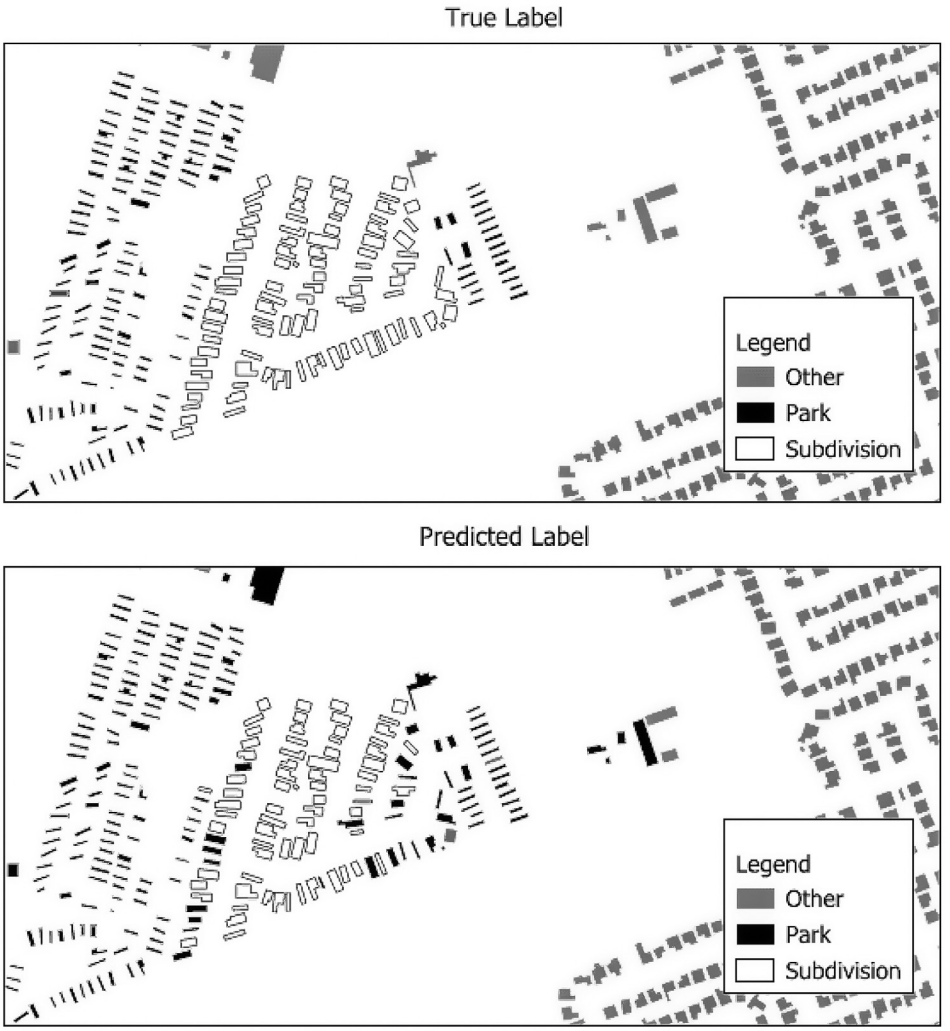


Sources: These data were produced by the authors. They are derived from analyses of building footprints acquired from Microsoft (2022) and OpenStreetMap (2023), parcel data from the Texas Natural Resources Information System Datahub, and the locations of manufactured home parks compiled by Durst et al. (2025) and Sullivan, Makarewicz, and Rumbach (2022)

Exhibit 4 illustrates the differences between the true and predicted labels for the same sample of building footprints shown in exhibit 2. The results suggest that the model performs well when identifying buildings in manufactured home parks (solid black) or other locations (solid grey) but has greater difficulty in identifying buildings in informal or manufactured home subdivisions. In the case of the manufactured home subdivision in exhibit 4, many of the buildings are mistakenly classified as manufactured home parks when, in fact, they are manufactured home subdivisions. This misclassification is primarily the case for buildings at the periphery of the subdivision and in close proximity to one of the two manufactured home parks.

Exhibit 4

True and Predicted Labels for Buildings in Parks, Subdivisions, and Other Locations



Sources: These data were produced by the authors. They are derived from analyses of building footprints acquired from Microsoft (2022) and OpenStreetMap (2023), parcel data from the Texas Natural Resources Information System Datahub, and the locations of manufactured home parks compiled by Durst et al. (2025) and Sullivan, Makarewicz, and Rumbach (2022)

Exhibit 5 presents performance metrics for manufactured home parks and informal or manufactured home subdivisions for the test dataset. Our models achieve a 99-percent accuracy for buildings in manufactured home parks and 95 percent for buildings in informal or manufactured home subdivisions. The results for precision illustrate that 70 and 46 percent of the buildings that our model predicts are manufactured home parks and informal or manufactured home subdivisions, respectively, and are classified correctly. The results for recall indicate that 87 and 63 percent of the buildings in manufactured home parks and informal or manufactured home subdivisions, respectively, are classified correctly by our model.

Exhibit 5

Performance Metrics

	Manufactured Home Parks			Informal or Manufactured Home Subdivision		
Validation	Test Dataset	DHS Site (0.1 Mile)	DHS Site (0.25 Mile)	Test Dataset	Pre-1990 Subdivisions	Pre-2010 Subdivisions
TP	1,455	517,224	1,020,970	3,229	7,445	14,535
FP	617	857,912	354,166	3,852	24,159	17,069
FN	210	1,256,160	5,218,141	1,909	78,477	96,734
TN	107,120	134,902,244	130,940,263	100,412	641,909	623,652
Precision	0.70	0.38	0.74	0.46	0.24	0.46
Recall	0.87	0.29	0.16	0.63	0.09	0.13
Accuracy	0.99	0.98	0.96	0.95	0.86	0.85

DHS = U.S. Department of Homeland Security. FN = false negative. FP = false positive. TN = true negative. TP = true positive.

Notes: This table presents measures of model performance. We compare our model predictions with the test dataset and with government or academic registries. We use distance (0.1 and 0.25 miles) to the nearest manufactured home community tracked by the DHS Homeland Infrastructure Foundation-Level Database to validate predictions for manufactured home parks. To validate predictions of informal or manufactured home subdivisions, we identify all buildings in the six Texas border counties with the largest number of colonias (Cameron, El Paso, Hidalgo, Maverick, Starr, and Webb). We then identify buildings within pre-1990 colonias monitored by the Texas Office of the Attorney General and post-1989 informal subdivisions database compiled by Durst (2016), which we use as validated subdivisions to evaluate the performance of our models.

Sources: These data were produced by the authors. They are derived from analyses of building footprints acquired from Microsoft (2022) and OpenStreetMap (2023), parcel data from the Texas Natural Resources Information System Datahub, and the locations of manufactured home parks compiled by Durst et al. (2025) and Sullivan, Makarewicz, and Rumbach (2022). Ancillary data sources include the location of manufactured home parks published by the DHS Homeland Infrastructure Foundation-Level Database, pre-1990 colonias from the Texas Office of the Attorney General, and post-1989 informal subdivisions compiled by Durst (2016)

To conclude our discussion of the data and methods, we examine the degree to which our predictions align with the coverage provided by existing registries of these communities. We first evaluate manufactured home park coverage by calculating the distance in miles between each of our buildings and the nearest manufactured home community documented by the DHS Homeland Infrastructure Foundation-Level Database. The DHS database provides only the latitude and longitude of the community, typically near a main road. Our models achieve an accuracy between 0.96 and 0.98 for buildings within 0.1 and 0.25 miles of a DHS location. This modest reduction in the estimated model performance is likely due to the DHS database including both manufactured home parks and manufactured home subdivisions and that not all buildings within the 0.1- and 0.25-mile buffer are necessarily within the manufactured home parks in question (Durst et al., 2025).

For informal or manufactured home subdivisions, we compare our estimates with validated registries documenting the location of these communities. We examine the six Texas border counties with the largest number of colonias (Cameron, El Paso, Hidalgo, Maverick, Starr, and Webb). We use the Texas Office of the Attorney General's colonias database to document the location of state-designated colonias developed before 1990, which we supplement with a database compiled by Durst (2016) that documents informal subdivisions developed in these same counties between 1990 and 2010. Across these six counties, we identify more than 750,000 buildings, 15 percent of which are in one of the two existing registries (Durst, 2016; Texas OAG, 2023). Exhibit 5 shows the model performance for these validated subdivisions. Our accuracy rates are 0.86 and 0.85 for pre-1990 and pre-2010 subdivisions, respectively—considerably less than the accuracy rate for the test dataset (0.95). This reduction in performance is likely due to the registries we use not containing any informal or manufactured subdivisions developed after 2010. Thus, these estimates of model performance are conservative.

Results

We identify more than 1.3 million buildings in parks and more than 2.2 million buildings in informal or manufactured home subdivisions nationwide (exhibit 6). We estimate that manufactured home parks account for approximately 1.1 percent of buildings in the Midwest and South, 0.8 percent in the Northeast, and 0.7 percent in the West, whereas buildings in informal or manufactured home subdivisions make up 0.9 percent of buildings in the Northeast, 1.1 percent of buildings in the Midwest and West, and 2.6 percent of buildings in the South. These percentages are likely underestimates of the true prevalence of these communities due to the relatively low recall of our models. To assess this potential underestimation by our models, we compare our regional estimates with Durst and Sullivan's (2019) findings from the American Housing Survey detailing the distribution of these neighborhoods across the four census regions (exhibit 7).[6] The relative similarity in the distribution across census regions suggests that both studies successfully capture geographic variation across the country. Both analyses suggest that most (61–75 percent) informal or manufactured home subdivisions are concentrated in the South, with only a fraction (4–8 percent) in the Northeast. Similarly, both studies find that a plurality (43–44 percent) of manufactured home parks are in the South, a substantial share (22–30 percent) in the Midwest, and relatively few (10–12 percent) in the Northeast.

[6] The counts from the two studies are not directly comparable because of differences in the universe—manufactured housing units in Durst and Sullivan (2019) versus building footprints in our study—but comparing the share of buildings in each census region is a useful robustness check.

Exhibit 6

Building, Locational, Socioeconomic, and Demographic Characteristics by Neighborhood Type

	Manufactured Home Parks	Informal or Manufactured Home Subdivisions	Other Locations
A. Building and Locational Characteristics			
Total Buildings	1,375,136	2,263,744	133,894,660
Singlewide Manufactured Homes (%)	55.2	15.2	0.2
Incorporated Cities (%)	50.8	40.7	53.0
Census Region			
% of All Buildings in the Midwest	1.1	1.1	97.8
% of All Buildings in the Northeast	0.8	0.9	98.3
% of All Buildings in the South	1.1	2.6	96.2
% of All Buildings in the West	0.7	1.1	98.2
B. Socioeconomic and Demographic Characteristics			
Population Density (People per Square Mile)	1,936	858	3,021
Non-Hispanic White (%)	65.7	68.2	68.8
Black (%)	9.9	11.3	10.2
Hispanic or Latino (%)	18.2	15.0	13.6
Median Household Income ($)	57,318	55,925	76,364
Poverty Rate	17.3	17.1	12.3
Unemployment Rate	6.0	6.1	5.4
Foreign Born (%)	8.9	6.2	8.9
Homeownership Rate	67.9	73.5	73.7
Median Housing Value ($)	162,184	149,295	271,014
Housing Vacancy Rate	11.9	15.4	12.3
Median Year Structure Built	1983	1983	1979

Notes: This table presents descriptive statistics for buildings in manufactured home parks, informal or manufactured home subdivisions, and buildings in other locations. Panel A presents building and locational characteristics. Panel B presents socioeconomic and demographic characteristics derived from 2018–2022 American Community Survey 5-year estimates acquired from the National Historical Geographic Information System (Manson et al., 2024). We measure these characteristics by identifying the census block group in which each building was located and calculate the average for each neighborhood type. We present statistics for a 1-percent sample of buildings in other locations.

Sources: These data were produced by the authors. They are derived from analyses of building footprints acquired from Microsoft (2022) and OpenStreetMap (2023), parcel data from the Texas Natural Resources Information System Datahub, and the locations of manufactured home parks compiled by Durst et al. (2025) and Sullivan, Makarewicz, and Rumbach (2022). Ancillary data were acquired from the 2018–2022 American Community Survey (Manson et al., 2024)

Exhibit 7

A Comparison of Regional Estimates

	The Authors	Durst and Sullivan (2019)
Informal or Manufactured Home Subdivisions		
Total Buildings/Units	2.3 million	2.1 million
Midwest (%)	16.9	6.3
Northeast (%)	7.9	3.9
South (%)	61.5	75.4
West (%)	13.7	14.5
Manufactured Home Parks		
Total Buildings/Units	1.3 million	2.7 million
Midwest (%)	29.6	21.5
Northeast (%)	11.9	10.4
South (%)	43.8	43.4
West (%)	14.7	24.8

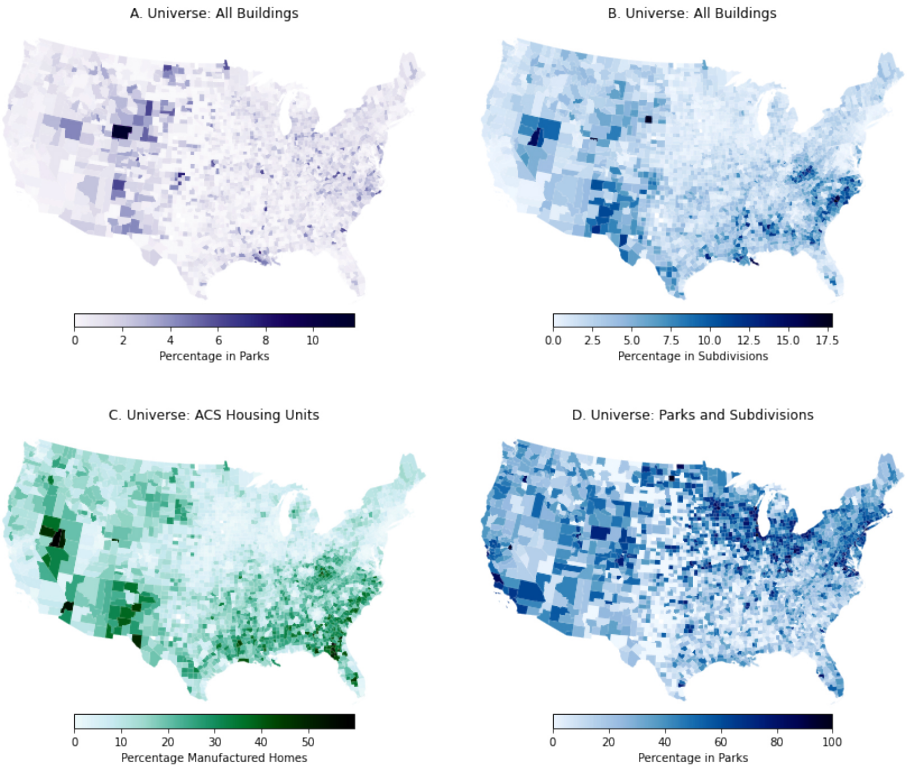
Notes: This table presents estimates of the distribution of manufactured home parks and informal and manufactured home subdivisions across census regions. We compare these estimates with regional estimates from Durst and Sullivan (2019) derived from the American Housing Survey. Durst and Sullivan (2019) estimate the total number of manufactured homes in each neighborhood, whereas we estimate the total number of buildings.

Sources: These data were produced by the authors. They are derived from analyses of building footprints acquired from Microsoft (2022) and OpenStreetMap (2023), parcel data from the Texas Natural Resources Information System Datahub, and the locations of manufactured home parks compiled by Durst et al. (2025) and Sullivan, Makarewicz, and Rumbach (2022). Ancillary data were acquired from Durst and Sullivan (2019)

In contrast to Durst and Sullivan's (2019) regional estimates, our method provides the first subregional estimates of the distribution of these neighborhoods. To illustrate, exhibit 8A plots county-level estimates of the percentage of buildings in manufactured home parks, and exhibit 8B shows estimates of the percentage in informal or manufactured home subdivisions. We find high concentrations of manufactured home parks in the Intermountain West, portions of Appalachia, and Louisiana. In contrast, we find high concentrations of informal or manufactured subdivisions in border counties throughout Texas and New Mexico, across the Intermountain West, and stretching across Southern states from Texas to Virginia. Exhibit 8C shows county-level variation in the share of housing units that are manufactured homes, as measured by 2018–2022 American Community Survey 5-year estimates, illustrating that manufactured homes are generally highly concentrated in Southern and Western states. Finally, exhibit 8D shows the relative concentration of parks. To do so, we compare the number of buildings in parks to the number in either parks or subdivisions at the county level. The results illustrate that parks are heavily concentrated in the Midwestern and Northeast states, whereas subdivisions are heavily concentrated in Southern states.

Exhibit 8

County-Level Estimates of the Prevalence of Manufactured Homes, Manufactured Home Parks, and Informal or Manufactured Home Subdivisions



ACS = American Community Survey.

Sources: These data were produced by the authors. They are derived from analyses of building footprints acquired from Microsoft (2022) and OpenStreetMap (2023), parcel data from the Texas Natural Resources Information System Datahub, and the locations of manufactured home parks compiled by Durst et al. (2025) and Sullivan, Makarewicz, and Rumbach (2022). Ancillary data were acquired from the 2018–2022 American Community Survey (Manson et al., 2024)

Our building footprint-derived estimates also provide insight into the characteristics of the broader communities in which these buildings are located. Exhibit 6 presents the average socioeconomic and demographic characteristics of block groups containing each building. We find that informal or manufactured home subdivisions are in areas with lower average population densities (858 people per square mile) and are less likely to be in incorporated cities (40.7 percent) than are manufactured home parks (1,936 people per square mile and 50.8 percent, respectively) and buildings in other locations (3,021 people per square mile and 53 percent, respectively). Thus, informal and manufactured home subdivisions are more rural and likely have lower levels of infrastructure and services than manufactured home parks or buildings in other locations. Although some differences exist in the socioeconomic and demographic characteristics of informal or manufactured home subdivisions and manufactured home parks, the most striking differences are between these two communities and buildings in other locations. For example, incomes are considerably lower in block groups that contain informal or manufactured home subdivisions ($57,318) or manufactured home parks ($55,925) than in other locations ($76,364). Respective

poverty rates (17 versus 12 percent) and housing values (less than $163,000 versus $271,000) also differ between parks or subdivisions and buildings in other locations. Notably, we do not find particularly large differences in the race and ethnicity or immigration status of residents in parks or subdivisions, suggesting that these communities provide affordable housing options to relatively diverse, although economically disadvantaged, residents.

Conclusion

Manufactured housing, predominantly in Southern and Western states, is one of the most affordable housing options today. Research has documented tenure insecurity and housing quality challenges faced by residents in manufactured home parks and informal subdivisions in the U.S.-Mexico border region, where manufactured housing is common (Esparza and Donelson, 2008; Sullivan, 2017, 2022; Ward, 1999). By comparison, research examining the prevalence or needs of informal or manufactured home subdivisions across the country is relatively limited (Durst and Sullivan, 2019; Reyes et al., 2024; Ward and Peters, 2007). In this study, we use building footprints and machine learning to develop the first subregional estimates of the prevalence and distribution of these communities across the continental United States. Our findings underscore the need for further research to understand the infrastructural and housing needs of these communities and how the distinct land tenure arrangements affect the housing stability and economic security of residents. HUD manages multiple funding sources, like the Community Development Block Grant program, and keeps records of pre-1990 colonias along the U.S.-Mexico border, including interactive maps for overlaying other data to assess community housing and infrastructural needs. Our building footprint-derived estimates might be used to expand the scope of federal efforts to track the locations of and conditions in these communities or to target funding to facilitate improvements to housing and infrastructure in them.

Acknowledgments

The authors thank Wisdom Henry, Shae Burnham, and Vivian Morales for their assistance with data collection. The National Science Foundation funded this study (award #2048562).

The datasets produced in this study and the scripts used to produce them are available at https://www.openicpsr.org/openicpsr/project/216681/version/V1/view. Interactive maps show the locations of parks and subdivisions nationwide at the census block group and county levels at https://arcg.is/1bCuqb1.

Authors

Noah J. Durst is an associate professor in the School of Planning, Design, and Construction at Michigan State University. Angelica Gacis is a Freelance Data Science Consultant. Nithya Mylakumar is a statistical analyst in the Department of Psychology at Michigan State University. Angela Perez is a Ph.D. candidate in the Department of Political Science at Michigan State University.

References

Biljecki, Filip, Yoong S. Chow, and Kay Lee. 2023. "Quality of Crowdsourced Geospatial Building Information: A Global Assessment of OpenStreetMap Attributes," *Building and Environment* 237: 110295. https://www.sciencedirect.com/science/article/pii/S0360132323003220.

Boehm, Thomas P., and Alan Schlottmann. 2006. "A Comparison of Household Mobility for Owned Manufactured, Traditional Owned, and Rental Units Using the American Housing Survey," *Journal of Housing Economics* 15 (2): 126–142. https://www.sciencedirect.com/science/article/abs/pii/S1051137706000143.

Durst, Noah J. 2016. "The Nature and Extent of Self-Help Housing in Texas: From Colonias to Model Subdivisions," *Journal of Planning Education and Research* 36 (2): 145–157. https://journals.sagepub.com/doi/10.1177/0739456X15612199.

Durst, Noah J., and Esther Sullivan. 2019. "The Contribution of Manufactured Housing to Affordable Housing in the United States: Assessing Variation Among Manufactured Housing Tenures and Community Types," *Housing Policy Debate* 29 (6): 880–898. https://www.tandfonline.com/doi/abs/10.1080/10511482.2019.1605534.

Durst, Noah J., Angelica Gacis, Nithya Mylakumar, Angela Perez, and Armin Yeganeh. 2025. "From Parks to Plots: Manufactured Housing in Different Contexts," *Urban Studies*. https://doi.org/10.1177/00420980251338753.

Durst, Noah J., Esther Sullivan, Huiqing Huang, and Hogeun Park. 2021. "Building Footprint-Derived Landscape Metrics for the Identification of Informal Subdivisions and Manufactured Home Communities: A Pilot Application in Hidalgo County, Texas," *Land Use Policy* 101. https://www.sciencedirect.com/science/article/abs/pii/S0264837720324960.

Esparza, Adrian X., and Angela J. Donelson. 2008. *Colonias in Arizona and New Mexico: Border Poverty and Community Development Solutions*. Tucson: University of Arizona Press. https://uapress.arizona.edu/book/colonias-in-arizona-and-new-mexico.

Fleischmann, Martin. 2019. "momepy: Urban Morphology Measuring Toolkit," *The Journal of Open Source Software* 4 (43): 1807. https://www.researchgate.net/publication/337406801_momepy_Urban_Morphology_Measuring_Toolkit.

Gonzales, Jack J. 2023. "Building-Level Comparison of Microsoft and Google Open Building Footprints Datasets." In *12th International Conference on Geographic Information Science*. Vol. 277. Leeds, UK: Leibniz International Proceedings in Informatics: 1–35. https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.GIScience.2023.35.

Manson, Steven, Jonathan Schroeder, David Van Riper, K. Knowles, Tracy Kugler, F. Roberts, and Steven Ruggles. 2024. "IPUMS National Historical Geographic Information System: Version 19.0 [dataset]." Minneapolis, MN: IPUMS. http://doi.org/10.18128/D050.V19.0.

Microsoft. 2022. "USBuildingFootprints." https://github.com/microsoft/USBuildingFootprints/.

Mukhija, Vinit, and David R. Mason. 2014. "Resident-Owned, Informal Mobile Home Communities in Rural California: The Case of Rancho Don Antonio, Coachella Valley," *Housing Policy Debate* 25 (1): 179–194. https://doi.org/10.1080/10511482.2014.921220.

Olmedo, Carlos, and Peter M. Ward. 2016. "Model Subdivisions: The New Face of Developer Lot Sales for Low-Income Colonia-Type Housing in Texas," *Land Use Policy* 15: 181–194. https://www.sciencedirect.com/science/article/abs/pii/S0264837715003944.

OpenStreetMap (OSM). 2023. https://planet.openstreetmap.org.

Pierce, Gregory, C.J. Gabbe, and Silvia R. Gonzalez. 2018. "Improperly-Zoned, Spatially Marginalized, and Poorly Served? An Analysis of Mobile Home Parks in Los Angeles County," *Land Use Policy* 76: 178–185. https://scholarcommons.scu.edu/ess/67/.

Reyes, Ariadna, Josh Newton, Camilo Caudillo, and Subham Kharel. 2024. "An Index to Identify and Classify the Spectrum of U.S. Peri-Urban Informal Subdivisions," *Journal of Planning Education and Research* 0 (0). https://doi.org/10.1177/0739456X241294232.

Rumbach, Andrew, Esther Sullivan, and Carrie Makarewicz. 2020. "Mobile Home Parks and Disasters: Understanding Risk to the Third Housing Type in the United States," *Natural Hazards Review* 21 (2): 05020001. https://www.researchgate.net/publication/338752557_Mobile_Home_Parks_and_Disasters_Understanding_Risk_to_the_Third_Housing_Type_in_the_United_States.

Sullivan, Esther. 2022. "Personal, Not Real: Manufactured Housing Insecurity, Real Property, and the Law," *Annual Review of Law and Social Science* 18 (1): 119–138. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4252396.

———. 2017. "Displaced in Place: Manufactured Housing, Mass Eviction, and the Paradox of State Intervention," *American Sociological Review* 82 (2): 243–269. https://journals.sagepub.com/doi/abs/10.1177/0003122416688667.

Sullivan, Esther, Carrie Makarewicz, and Andrew Rumbach. 2022. "Affordable but Marginalized," *Journal of the American Planning Association* 88 (2): 232–244. https://ideas.repec.org/a/taf/rjpaxx/v88y2022i2p232-244.html.

Texas Office of the Attorney General (Texas OAG). 2023. "Texas Colonias Viewer." https://texasoag.maps.arcgis.com/apps/webappviewer/index.html?id=1bc9c4f7b1da47dd8fc535fbd17dc060.

U.S. Department of Homeland Security (DHS). 2023. "Homeland Infrastructure Foundation-Level Data - Mobile Home Parks Layer." https://hifld-geoplatform.hub.arcgis.com/datasets/mobile-home-parks/about.

Ward, Peter M. 1999. *Colonias and Public Policy in Texas and Mexico: Urbanization by Stealth*. Austin: University of Texas Press. https://utpress.utexas.edu/9780292791251/.

Ward, Peter M., and Paul A. Peters. 2007. "Self-Help Housing and Informal Homesteading in Peri-Urban America: Settlement Identification Using Digital Imagery and GIS," *Habitat International* 31 (2): 205–218. https://www.researchgate.net/publication/222559507_Self-Help_Housing_and_Informal_Homesteading_in_Peri-Urban_America_Settlement_Identification_Using_Digital_Imagery_and_GIS.