

Updates to USA Structures (Phase 3 Prototype)

2023–2024 Technical Report



H. Lexie Yang
Taylor Hauser
Benjamin Swan
Joe Tuccillo
Abhishek Potnis
Daniel Adams
Jessica Moehl
Joshua Mcateer

**Approved for public release.
Distribution is unlimited.**

October 2024



DOCUMENT AVAILABILITY

Online Access: US Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free via <https://www.osti.gov/>.

The public may also search the National Technical Information Service's [National Technical Reports Library \(NTRL\)](#) for reports not available in digital format.

DOE and DOE contractors should contact DOE's Office of Scientific and Technical Information (OSTI) for reports not currently available in digital format:

US Department of Energy
Office of Scientific and Technical Information
PO Box 62
Oak Ridge, TN 37831-0062

Telephone: (865) 576-8401

Fax: (865) 576-5728

Email: reports@osti.gov

Website: <https://www.osti.gov/>

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Geospatial Science and Human Security Division

**UPDATES TO USA STRUCTURES (PHASE 3 PROTOTYPE)
2023–2024 TECHNICAL REPORT**

H. Lexie Yang
Taylor Hauser
Benjamin Swan
Joe Tuccillo
Abhishek Potnis
Daniel Adams
Jessica Moehl
Joshua Mcateer

October 14, 2024

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831
managed by
UT-BATTELLE LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

CONTENTS

LIST OF FIGURES	iv
LIST OF TABLES	v
LIST OF ABBREVIATIONS	vi
1. UPDATES TO STRUCTURE FEATURES	1
1.1 IMAGERY CURATION	1
1.2 MODEL DEVELOPMENT AND DEPLOYMENT	2
1.3 PERFORMANCE EVALUATION FOR USA STRUCTURES UPDATE	4
2. MODERNIZATION OF THE POSTPROCESSING WORKFLOW	6
2.1 GEOMETRIC REGULARIZATION	7
2.2 MERGE GEOMETRY UPDATES WITH EXISTING USA STRUCTURES DATA	7
3. NEW ATTRIBUTION OF BUILDING CODE (B_CODE)	10
3.1 SOURCE DATA	10
3.2 DATA TRANSFORMATION	10
3.3 DATA CONFLATION	11
4. ESTIMATION OF POPULATION COUNTS	13
5. FEEDBACK FROM THE US GEOLOGICAL SURVEY NATIONAL MAP CORPS	15
6. ONGOING WORK	17
6.1 CHANGE DETECTION WITH SYNTHETIC APERTURE RADAR IMAGES	17
6.2 ResType MODEL TRAINING	19
6.3 IMPROVEMENT TO AUTOMATIC QA/QC	22
7. REFERENCES	24

LIST OF FIGURES

Figure 1.	Image coverage of the nine priority states.	1
Figure 2.	Comparisons between raw geometries from Phase 1 (in blue) and Phase 3 (in red). . .	3
Figure 3.	Illustration of how areas were selected for manual QA/QC.	5
Figure 4.	Phases 1, 2, and 3 postprocessing workflows.	6
Figure 5.	Phase 1 and Phase 3 regularzation workflows.	7
Figure 6.	Regularization results in order of computational expense.	8
Figure 7.	Comparisons between new polygons and existing polygons.	9
Figure 8.	Frequency distribution of the B_CODE attribute by year and state.	12
Figure 9.	USA Structures in Florida symbolized by the B_CODE Attribute.	12
Figure 10.	Preliminary results of change detection workflow.	17
Figure 11.	Limitations of change detection workflow.	18
Figure 12.	Multivariate and sensitivity testing results for the study area.	22
Figure 13.	State-by-state sensitivity testing results.	23

LIST OF TABLES

Table 1.	State-level breakdown of validation results.	4
Table 2.	Categorical breakdown of common causes of false positive building detections. . . .	5
Table 3.	USA Structures development phases.	6
Table 4.	BCAT field names and descriptions.	10
Table 5.	Modified BCAT field names and descriptions.	11
Table 6.	Conflation criteria between USA Structures and UrbanPop by dwelling type.	13
Table 7.	Summary of USGS feedback and edits to USA Structures data over St. John the Baptist Parish, Louisiana.	15
Table 8.	Bayesian optimization results for various sampling techniques across regions.	21
Table 9.	Summary of state-by-state multivariate tests.	23

LIST OF ABBREVIATIONS

ACS	American Community Survey
ADASYN	Adaptive Synthetic Sampling Approach
AFE	automatic feature extraction
BCAT	Building Code Adoption Portal
FEMA	Federal Emergency Management Agency
FP	false positive
GAUNTLET	Geographic Augmentation of Extracted Building Features Tool
ORNL	Oak Ridge National Laboratory
QA	quality assurance
QC	quality control
SAR	synthetic aperture radar
SMOTE	Synthetic Minority Oversampling Technique
USGS	US Geological Survey
VVM	verification and validation model

EXECUTIVE SUMMARY

The Oak Ridge National Laboratory (ORNL) team has successfully innovated geospatial data feature mapping and postprocessing workflows to support the USA Structures project supported by FEMA. This technical report highlights significant updates, improvements, and innovations for the USA Structures dataset, focusing on nine priority states. Key accomplishments include:

- Deployment of a transformer-based AI model for advanced structure mapping
- Integration of FEMA’s Building Code Adoption Portal data for new B_CODE attribution
- Utilization of ORNL’s UrbanPop synthetic population model for population estimation
- Modernization of post-processing workflow for larger-scale data refreshes
- Investigation of Sentinel-1 synthetic aperture radar imagery for change detection in built environments
- Research outcomes on class imbalance management for ResType model enhancement

This report details the technical strategies developed. Building on the success of this prototyping stage covering nine priority states, we anticipate continuing the R&D activities and scaling these processes to the remaining states. These ongoing initiatives underscore our commitment to continuous improvement and technological leadership in supporting FEMA’s mission.

1. UPDATES TO STRUCTURE FEATURES

The current USA Structures dataset was based on the structure features created and completed 2022 (Phase 1 hereafter). The satellite images used to map those structure features are dated between 2011 and 2021, so they may be outdated and can create discrepancies that could cause failure to capture recent structural changes in the environment. Furthermore, the latest advances in computer vision for mapping structures from recent high-resolution remote sensing images will help improve the geometries for structural features. This update involved curating images covering nine priority states (i.e., Texas, Louisiana, Mississippi, Alabama, Georgia, Florida, South Carolina, North Carolina, Virginia) and deploying the latest automatic feature extraction model based on the **OREOLE** foundational model recently developed by Oak Ridge National Laboratory (Dias et al. 2024). The updated structure polygons will be used to attach attributions that were included in Phase 2 (Hauser et al. 2023). Two new attributions are also added in this update, denoted as Phase 3 hereafter.

1.1 IMAGERY CURATION

In this update, satellite images were still considered as the source of data for mapping structures as previous phases. To capture the most recent changes from imagery and ensure the foundational structure data layer is up to date, 5,831 WorldView-2 and WorldView-3 images were curated. These images are newer than the images previously used for mapping existing polygons in the dataset. These new images were collected between 2017 and 2024. The preprocessing steps were the same as those described in a previous Data Descriptor (Yang et al. 2024). However, image coverage gaps in the nine states, shown in Figure 1, may have been caused by the image curation criteria (e.g., near-nadir looking angle, minimal cloud coverage).

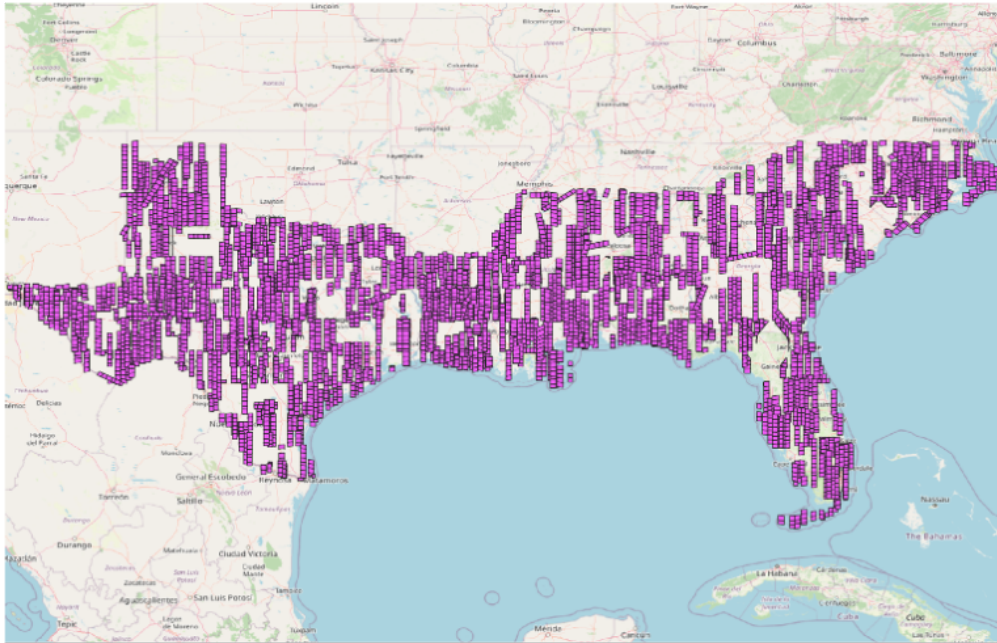


Figure 1. Image coverage of the nine priority states.

1.2 MODEL DEVELOPMENT AND DEPLOYMENT

The geometries from Phase 1 were created largely based on computer vision models and training data that were developed before 2021. To provide better raw geometries from computer vision models while avoiding the need to acquire new labeled samples, this project leveraged recent work (Dias et al. 2024) in which a transformer-based foundation model was derived from a collection of high-resolution satellite imagery to improve model generalization and structure geometries.

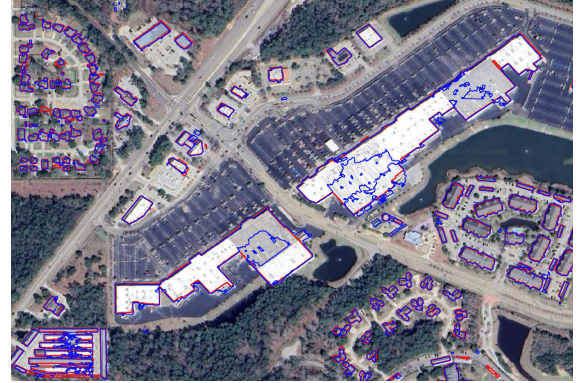


Figure 2. Comparisons between raw geometries from Phase 1 (in blue) and Phase 3 (in red).

The foundation model was further fine-tuned using the same internally developed, large-scale building training sample set (Swan et al. 2024). The fine-tuned model was then deployed to process those 5,831 new images. With a larger and more advanced architecture, it has achieved better performance in precision, recall, and visual fidelity to building features in images compared with previous iterations. The two examples in Figure 2 show that the new model yields more complete and clean geometries for larger buildings compared with the raw geometries from Phase 1.

1.3 PERFORMANCE EVALUATION FOR USA STRUCTURES UPDATE

Because the transformer model had not previously been deployed for structure mapping at scale, accuracy measures specific to this project area were not available. Therefore, this work also involved producing some for project reference and to better inform users of the rough expectations of the data quality. To create the accuracy measures, a manual quality assurance (QA) and quality control (QC) was performed on statistical hot spots of newly detected structures, creating a series of contiguous review zones where all structures were evaluated to eliminate false positive (FP) detections. By focusing on hot spots, the impact of manual quality review could be maximized within minimal time. The procedure for obtaining these hot spots was to map new detections to a 5 km grid and perform a Getis–Ord G_i^* cluster analysis (Getis and Ord 1992), taking the grid cells with a z-score of 2 or greater as the QA/QC zones. An example of gridded new detections and cluster analysis for Georgia is shown in Figure 3.

As given in Table 1, a review of 1.3 million detections over an area of 65,000 km² yielded 7,723 FP detections, for a precision of 99.43%. Because of time constraints, the review of the of hot spot zones in Florida was not fully systematic; therefore, those zones were not included in this evaluation set for the purposes of obtaining a validation metric.

The manual review process involved trained reviewers assessing each detected structure against specific criteria, including size, shape, and context within the landscape. Findings from this assessment were documented to guide future refinements of the detection algorithms based on any misclassifications.

Throughout the review, several features, such as airports, shipping containers, and trains, were consistently responsible for most FP predictions. A categorized summary is presented in Table 2. This information can help improve workflow and algorithms (e.g., including those FPs as negative training set) because it can guide future QA/QC efforts and sample generation for additional training of building extraction models and automated QA/QC models.

Table 1. State-level breakdown of validation results.

State	Area reviewed (km2)	New building detections	False positives	Precision
Alabama	11,325	107,132	558	99.48%
Georgia	8,300	177,844	851	99.52%
Louisiana	5,725	46,679	593	98.73%
Mississippi	4,675	60,092	263	99.56%
North Carolina	6,350	145,460	1,207	99.17%
South Carolina	6,000	156,235	520	99.67%
Texas	18,700	314,284	2,383	99.24%
Virginia	3,950	336,431	1,348	99.60%
Total	65,025	1,344,157	7,723	99.43%

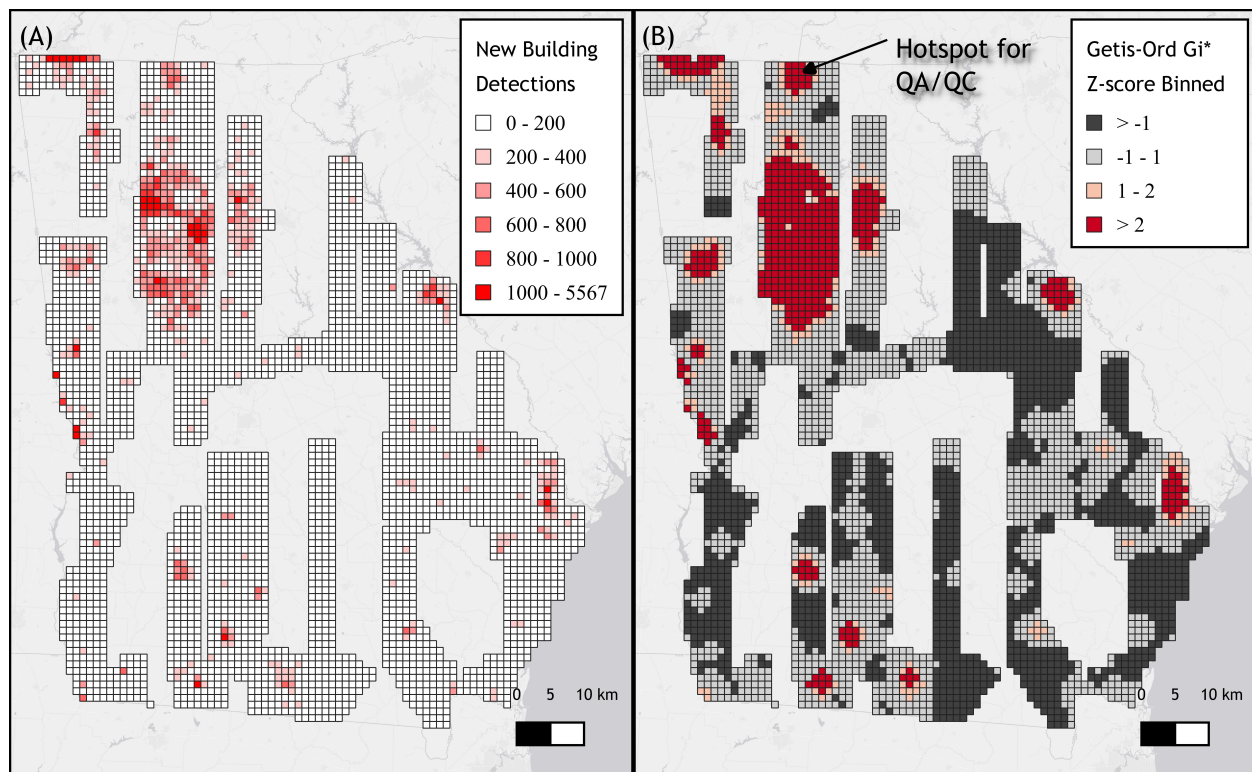


Figure 3. Illustration of how areas were selected for manual QA/QC. (A) New building detections in Georgia mapped as a 5 km grid. (B) Getis–Ord G_i^* analysis with z-scores binned for display. Grid cells with a z-scores of 2 or greater were chosen for review.

Table 2. Categorical breakdown of common causes of false positive building detections.

Category	Examples
Temporary structures	Construction sites, mobile office trailers
Building-like features	Shipping containers, recreational vehicles, train cars
Miscellaneous manufactured features	Roads/highways (especially overpasses), boat-s/barges, helipads, athletic fields/courts, bleachers, solar panels, transmission lines, air/seaport-s/marinas, billboards
Natural features	Open water, beaches/sandbars, off-coast white-caps

2. MODERNIZATION OF THE POSTPROCESSING WORKFLOW

The overall workflow for USA Structures is highly complex and requires broad expertise, including remote sensing, to geospatial data science, vector data engineering, and modeling. This section discusses the changes over time to the postprocessing portion of the workflow. The postprocessing portion starts as soon as the raw (unregularized and unvalidated) structure detections are extracted from rasters and vectorized into spatial vector polygons.

During the entire development of USA Structures (since 2017), technologies and goals have changed. USA Structures has undergone three major phases of development. Table 3 describes postprocessing of these three phases and their focus. Figure 4 shows the workflow of all three phases side by side. Phase 1 focused on development of final schema, QA/QC (automated and manual), regularization, and sources for attributes. Phase 1 technology's stack was Python 2.7 and ArcPy. All postprocessing was done on features classes within file geodatabases. Phase 2's processing environment was PostGIS 14, Python 3.5, and docker containers. Phase 3 sought to resolve the differences between Phase 1 and Phase 2 while improving regularization output from an Environmental Systems Research Institute, Inc. (ESRI)—specific tool. Phase 3 transferred all ArcPy-required steps of Phases 1 and 2 into PostGIS, Python 3.10, and docker containers, with the exception of one step within the regularization portion of the workflow. Figure 5 illustrates the differences between Phase 1 and Phase 3 regularization workflows. Another Phase 3 innovation was a workflow built so that manual QA and the vast majority of the workflow could be executed in parallel. This configuration provides more time for manual QA while removing it as a potential bottleneck for computationally expensive tasks that usually follow the manual QA step.

Table 3. USA Structures development phases.

Phase	Year(s)	Postprocessing focus
1	2017–2021	Establish baseline structure inventory and standard attribution
2	2021–2023	Addresses, OCC_CLS, and PRIM_OCC
3	2023–2024	Structure updates, B_CODE, population fields, USGS feedback, and workflow modernization

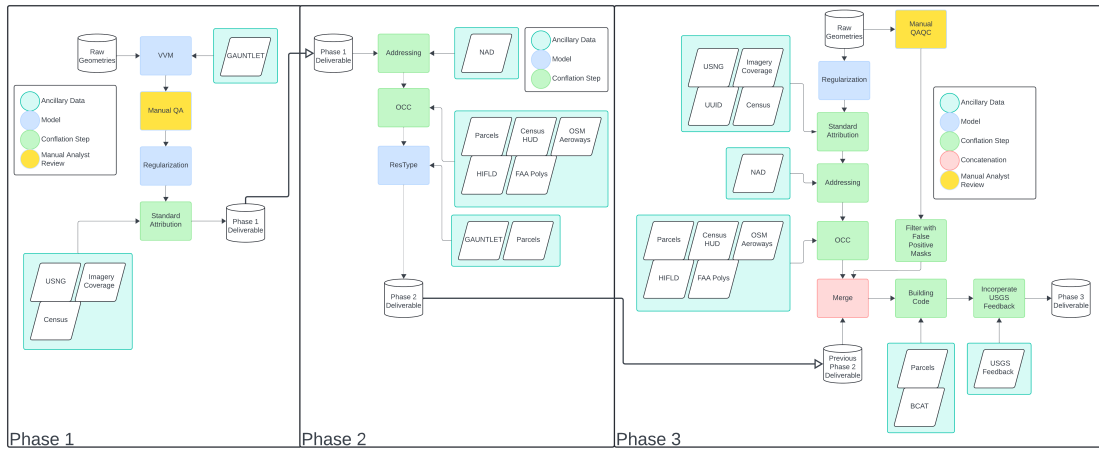


Figure 4. Phases 1, 2, and 3 postprocessing workflows.

2.1 GEOMETRIC REGULARIZATION

Following the new process (Figure 5), the raw geometries underwent a regularization process. After the geometric regularization step, a new set of structure polygons are ready for attaching attributions. One of the goals of this year is to improve the structure geometries of the dataset. Rather than reprocessing existing structure polygons, this need was addressed by two steps. First, the performance of the structure mapping model was improved, which is outlined in the Section 1. Second, extensive geometry simplification parameters were tested. The details of geometric regularization parameters are discussed below.

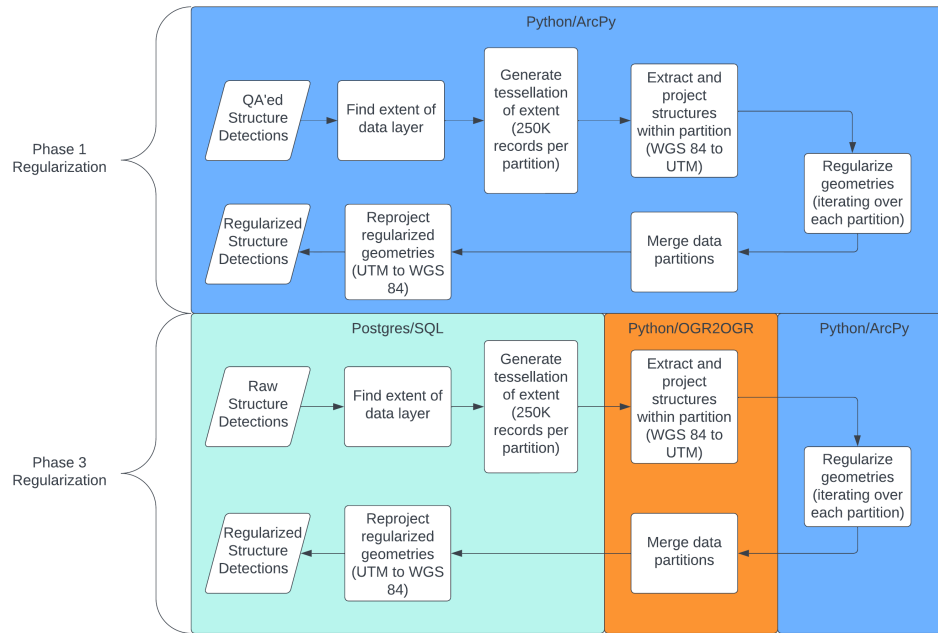


Figure 5. Phase 1 and Phase 3 regularization workflows.

Figure 6 illustrates the various outcomes for regularization using different parameters. Panel A shows raw geometries without any regularization. Panels B–D show the output from increasingly more precise parameters. The more precise the parameters, the more computation time is required. Phase 3 utilizes the set of parameters illustrated in Panel C for regularization. This level of precision greatly improved the raw geometries visually and improved on the Phase 1 regularization output. This change in parameters increased processing time by 30 min for every 1.25 million records. Previously, 1.25 million records took roughly 5 h of processing time; now the improved output takes 5.5 h.

Overall, using these two approaches together led to significant quality improvements in new geometries.

2.2 MERGE GEOMETRY UPDATES WITH EXISTING USA STRUCTURES DATA

With the new set of structure polygons provided by the transformer-based structure mapping model, several data-merging strategies were considered to leverage both existing USA Structure polygons and new updates. Based on comparisons among different merging approaches, the following priorities were adapted to select geometries to be included in the resulting updated dataset: (1) lidar-derived polygons, (2) newly satellite-derived and geometrically regularized polygons, and (3) existing structure polygons in the database. Lidar-derived structures were prioritized over all other sources because of a few key observations. The quality of satellite images of derived structures is affected by facades and shadows. This limitation stems

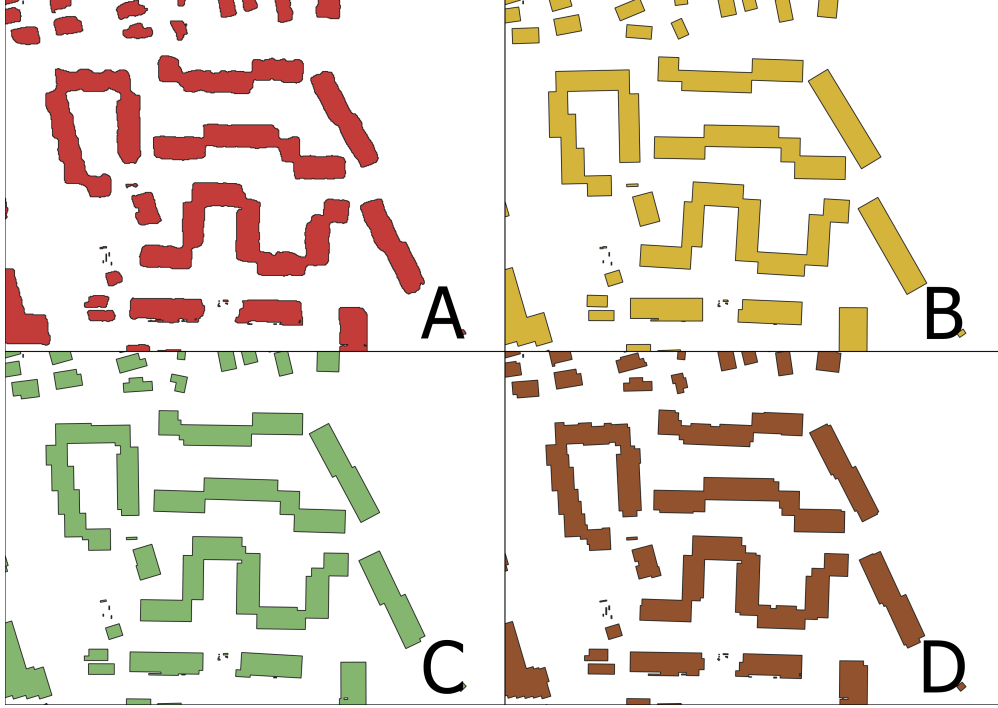


Figure 6. Regularization results in order of computational expense. (A) Raw geometries, (B) less precise parameters, (C) fairly precise parameters, and (D) more precise parameters.

from several factors inherent to satellite-based sensing. First, the oblique viewing angle of many satellite sensors can lead to significant geometric distortions, especially for tall structures. This effect causes buildings to appear to lean away from the nadir point, making their true ground footprint difficult to determine. It also causes the facade being identified as part of structures for those high rises in off-nadir images. Additionally, shadows cast by skyscrapers can obscure neighboring structures or create false impressions of building outlines. The lidar-derived structures might be outdated, but they are free from those drawbacks. Fortunately, lidar data coverage is minimal for the nine priority states being updated. Therefore, we chose lidar as the prioritized data source in the merging process.

The order of the next two sources is determined based on our visual comparisons, as shown below. The detected new structures are more complete and with clearer outlines/boundaries (as compare to the reference images).

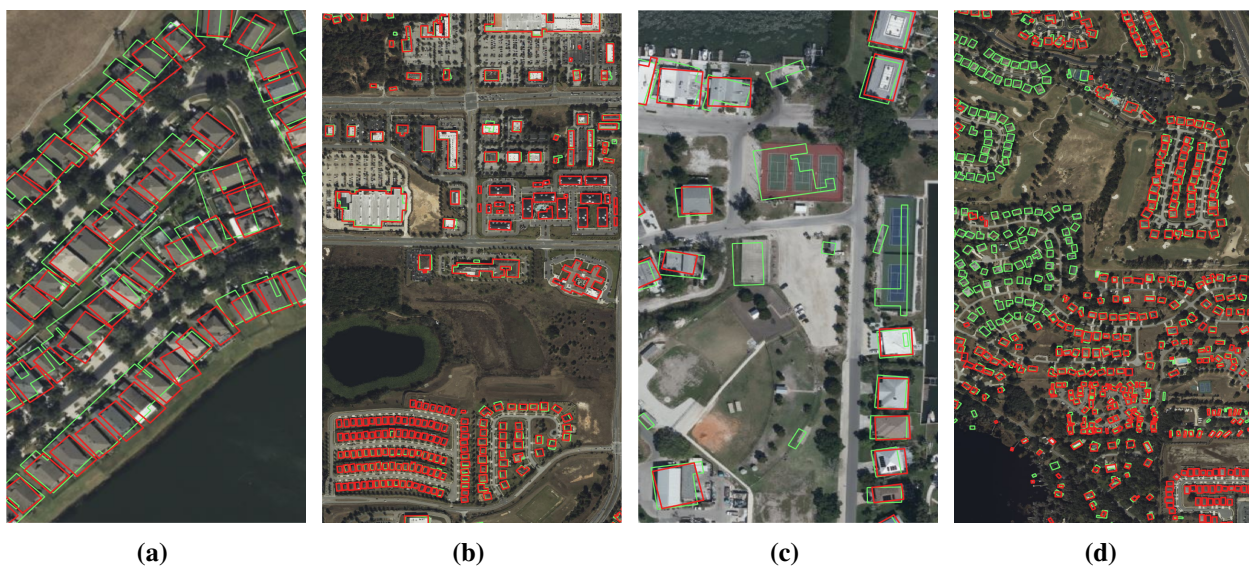


Figure 7. Comparisons between new polygons and existing polygons. Red boxes denote newly mapped structures from latest images, and green boxes indicate existing structures in the current USA Structures database. (a) More individual structures extracted, (b) new structures mapped with newer images, (c) fewer false positives, (d) new and existing structures. Clouds prevented new structures from being mapped on the left of this example.

3. NEW ATTRIBUTION OF BUILDING CODE (B_CODE)

The attribute B_CODE describes the likely standard building code that was in effect during the building's construction. This section describes the original source data, its modification, and the conflation methodologies to create the B_CODE attribute.

3.1 SOURCE DATA

Oak Ridge National Laboratory (ORNL) used Federal Emergency Management Agency (FEMA)'s Building Code Adoption Portal (BCAT) data to populate the B_CODE field. BCAT tracks building code adoption from state to local governments across the United States. (FEMA, [n.d.](#)). ORNL received the data from FEMA's Building Science division in the form of a comma-separated values file. The dataset included all states with mandatory statewide building codes and the time of adoption. The dataset is dated to 2020 with quarterly updates as recent as 2024. The temporal range of building codes is from 1914 to 2024, and more than 140 building codes are included. Table 4 describes the original columns of the BCAT data ORNL received.

Table 4. BCAT field names and descriptions.

Field name	Description
state	State abbreviation
version	State name of the building code standard
model_building_code	Standard commercial code equivalent
effective_start_date	Starting enforcement date of commercial building code
end_date	Retirement date of commercial building code
model_residential_code	Standard residential code equivalent
effective_start_date_2	Starting enforcement date of residential building code
end_date_2	Retirement date of residential building code

3.2 DATA TRANSFORMATION

The first modification to the original source data involved reorganizing the overall shape of the table. If the row of the table had information about a commercial building code, then the columns pertaining to the residential codes were left null. The opposite was true when the row had information concerning commercial codes. This process led to a wide table with the appearance of sparse attribution. ORNL added an `agg_type` field, which represents residential or nonresidential codes. Afterward, ORNL condensed `model_building_code`, `effective_start_date`, `end_date`, `model_residential_code`, `effective_start_date_2`, and `end_date_2` into three fields: `model_building_code`, `effective_start_date`, and `end_date`.

The second modification was to alter the overlap between a retired code and a new code within the same year. All new codes were shifted to the following year to generate a more conservative estimate of building codes. For example, if Code A ended on June 16, 2020, and Code B started on June 17, 2020, then the resulting `end_date` for Code A would be 2020, and Code B's `effective_start_date` would be 2021. Table 5 describes the modified table.

Table 5. Modified BCAT field names and descriptions.

Field name	Description
state	State abbreviation
code_version	State name of the building code standard
model_building_code	Standard code equivalent
effective_start_date	Starting enforcement date of building code
end_date	Retirement date of building code
mod_start_year	Rounded year in which the code enforcement began
mod_end_year	Rounded year in which the code enforcement ended
agg_type	Residential or Nonresidential aggregation

3.3 DATA CONFLATION

The conflation process uses a structure's location, OCC_CLS attribute, and intersecting parcel to assign a model_building_code from Table 5 to each structure. This process can result in a null value. Nulls are assigned when a structure's parcel has a null value for the year_built attribute, or if the year built precedes the beginning of the building code lineage. For example, a residential structure was built in 1962, but the first residential code's effective start date is 1968. Because the depth of lineage from state to state can vary, the completeness of the B_CODE attribute may vary greatly between states. Some states, such as Texas, do not have a statewide mandated building code adoption. In these instances, no building codes are assigned within the state.

The conflation workflow is executed as follows:

- Aggregate a structure's OCC_CLS into either residential or nonresidential categories.
- Use the parcel that contains the structure year_built attribute.
- Assign the proper building code that corresponds to the correct year and occupancy aggregation.

The frequency distribution of building codes per state is described in Figure 8. Figure 9 shows an example of output in Florida. Four states (Florida, Georgia, North Carolina, Virginia) have coverage of building codes greater than 25%, but only one state had coverage greater than 50% (Florida). If ORNL could obtain a longer lineage of residential codes for all states, then coverage of this important attribute could be greatly increased. Unfortunately, the limitation of this methodology is highlighted by the Texas results. Texas has 0% coverage because it lacks statewide mandated building codes. The effort to acquire local-level building code lineages would be significant. Figure 9 shows how this information can vary significantly from neighborhood to neighborhood.

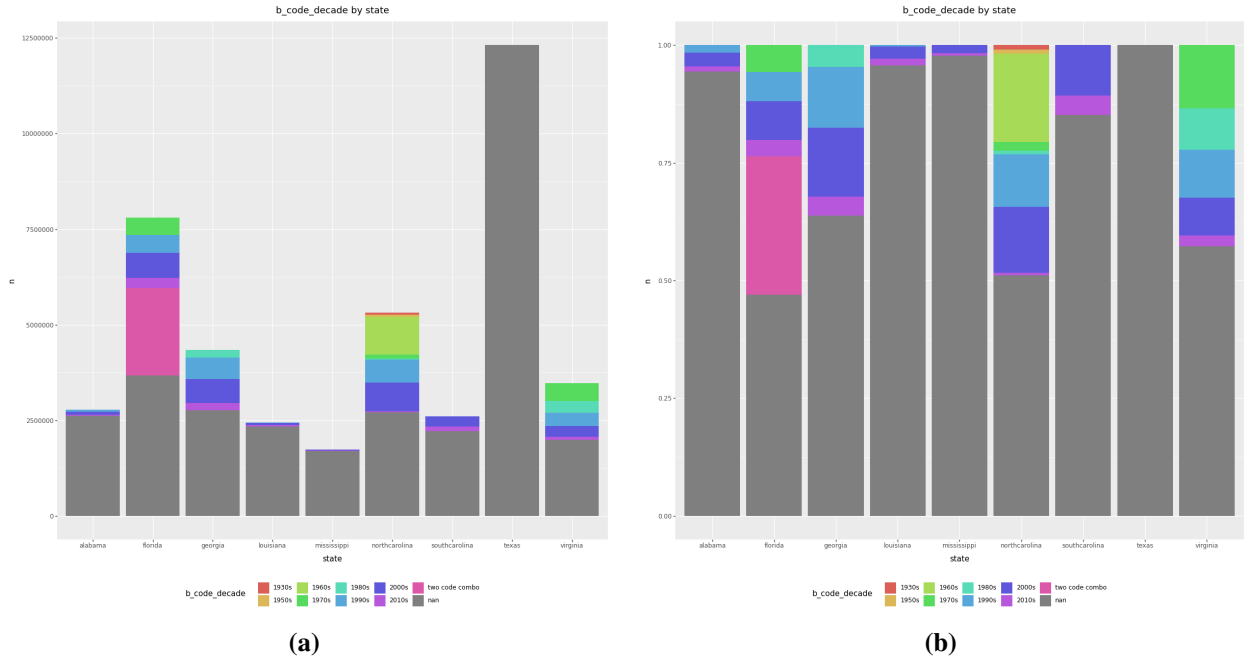


Figure 8. Frequency distribution of the B_CODE attribute by year and state. (a) Absolute value of records by state assigned a building code from specific decade. (b) A percentage of records within a state assigned a building code from a specific decade.



Figure 9. USA Structures in Florida symbolized by the B_CODE Attribute.

4. ESTIMATION OF POPULATION COUNTS

Population ranges were estimated for all residential structures in the nine priority states, except those with a `prim_occ` label of “Temporary Lodging.” This procedure leverages capabilities from ORNL’s UrbanPop synthetic population model (Tuccillo et al. 2023) to produce custom estimates of household size by dwelling type for census block groups, small enumeration areas with populations of roughly 600–3,000 residents. UrbanPop provides a framework for generating synthetic populations from the US Census Bureau’s American Community Survey (ACS) via bootstrap sampling of anonymized household records from its public use microdata sample to census block groups. This process faithfully preserves ACS summary file statistics on demographics, socioeconomic status, housing, and mobility published about each block group in the area of interest (Tuccillo et al. 2023). Each synthetic residence is attributed by dwelling type (single-family detached, single-family attached, multifamily 2–50+ units, mobile homes, group quarters), household income, and number of people.

Table 6. Conflation criteria between USA Structures and UrbanPop by dwelling type.

Dwelling type	ACS label	Expected unit count	Conflation variable: household	Conflation variable: USA Structures
Single-family detached	Single-family detached	1	Household income	Building floor area
Mobile home	Mobile home	1	Household income	Building floor area
	Other	1	Household income	Building floor area
Multifamily	Single-family attached	6	Units in structure	Units in structure
	2 units	2	Units in Structure	Units in Structure
	3–4 units	3.5	Units in structure	Units in structure
	5–9 units	7	Units in structure	Units in structure
	10–19 units	14.5	Units in Structure	Units in structure
	50 or more units	125	Units in structure	Units in structure

Following one process (Tuccillo 2024), UrbanPop dwelling characteristics were conflated with USA Structures as

$$\Gamma = \frac{1}{\exp(|h - s|)} \mid s \in S, \quad (1)$$

where h is the UrbanPop household conflation variable, and s is the USA Structures conflation variable. The choice of conflation variables depends on dwelling type (S) (Table 6). For one-unit dwellings such as single-family detached structures or mobile homes¹, building floor area was conflated with ranked household income. For multifamily structures, an expected number of units in the synthetic household’s dwelling (computed as the median of the labeled unit range) was conflated with the estimated number of units for the structure. In both cases, building floor area was computed as

$$BFA = A \times F, \quad F = \frac{H - 2}{3}, \quad (2)$$

1. “Other” dwelling types, labeled in the ACS as “Boat, RV, Van, etc.” were associated with mobile homes because they are assumed to be the closest analogue among the available USA Structures labels.

where A represents a structure’s polygon area, and F represents the number of floors in the structure as a function of its height in meters, H . Where lidar-derived heights were unavailable, heights were predicted from morphological properties of structures via ORNL’s Vivaldi model (Stipek et al. 2024).² Unit counts were estimated for multifamily structures using estimates of people per 1,000 ft² from ORNL’s Population Density Tables project (Stewart et al. 2016; Urban et al. 2023), specifically the median nighttime estimate for multifamily residential urban facilities in the United States (2.14 people/1,000 ft²).

Conflation scores were computed for residential structures and synthetic households within each block group relative to dwelling type labels (Eq. [1]). The conflation scores were in turn used to generate a weighted distribution of household sizes for each structure, which was then converted to a population distribution. For multiunit dwellings, household counts were multiplied by the number of units in the structure, whereas for single-unit dwellings, the population distribution was considered equivalent to that of households. Finally, the population ranges were summarized for each structure, providing a 95% coverage interval (percentiles 2.5 and 97.5) and median (percentile 50) population counts.

Group quarters (college dormitories, nursing homes, prisons) are a special type of residential facility featuring group living conditions rather than discrete housing units.³ In these cases, a fixed population was estimated for the structure: $BFA \times PDT$, where PDT is set as either (1) the multifamily residential urban facility estimate for the United States of 2.14 people/ft² for nursing homes (`prim_occ == "Nursing Home"`) or (2) the prisons/jails facility estimate of 2.09 people/ft² for institutional dormitories (`prim_occ == "Institutional Dormitory"`).⁴

The USA Structures population estimates, aggregated by state and county, were validated using published population totals obtained at those scales from the ACS (variable `B001001_001E`). Specifically, the degree to which the 95% coverage interval of the USA Structures estimates contained the ACS totals at each scale was evaluated. For states, all 95% coverage intervals for USA Structures contained the ACS population totals. For counties, slightly less correspondence was observed: 98.3% of the 95% coverage interval estimates for USA Structures contained the ACS population totals. These discrepancies are likely due to population estimates being performed on a per-structure basis. In future releases, place-specific conditions that may underlie these differences will be examined.

2. Height predictions are currently available for the USA Structures Phase 2 internal release only. As a workaround, predicted heights were averaged by block group and applied to the corresponding Phase 3 structures, using county and state-level averages as fail-safes when block group level labels were inconsistent between the two releases.

3. <https://www.census.gov/newsroom/blogs/random-samplings/2021/03/2020-census-group-quarters.html>

4. To date, college dormitories have been labeled in USA Structures releases as `occ_cls == "Education"` and therefore are not currently considered residential. This detail will be refined in future USA Structures releases, which will in turn require revised rules for assigning occupancy estimates to institutional dormitories.

5. FEEDBACK FROM THE US GEOLOGICAL SURVEY NATIONAL MAP CORPS

ORNL received and evaluated feedback from the US Geological Survey (USGS) National Map Corps as part of data quality and maintenance initiatives for improving the USA Structures dataset. The feedback was obtained for the areas of interest over St. John the Baptist Parish, Louisiana. This feedback aimed to validate or correct the occupancy types taken from the Phase 2 USA Structures data. These evaluations focused on examining the changes in records, providing statistics on the types of those changes, and noting ORNL observations.

The data were provided in a compressed file geodatabase format. All field names were converted to lowercase, and two new fields were added: “edited” and “last_edited_userid.” Out of 12,328 records, 11,915 (96.6%) had a value of –1 in the edited field. All records with a value of –1 in the edited field also had a username in the last_edited_userid field. The data were presented in the WGS 84 projection, which is the standard projection for USA Structures deliverables.

Analysis of the data revealed that 9,305 (75.5%) structures had attribution that remained unchanged. If each building was validated, then this number is encouraging. That unchanged attribution was interpreted as a verification of the building’s occupancy classification when the edited field had a value of –1. This value indicates that the analyst either agrees with ORNL’s classification or could not find enough information to change the classification. A total of 3,023 (24.5%) of structures had their attribution modified, and 35% (rows 1 and 2 in Table 7) of all changes involved changing a residential structure from single family to manufactured home or vice versa. FEMA’s mission could be better supported by including such validated and corrected information in the dataset updates because manufactured homes are vulnerable to natural disasters. Additionally, building code information will be added to those corrected single-family dwelling records.

Of all attribution changes, 8.3% (Rows 4 and 5 in Table 7) were industrial structures changed from light to metals/minerals/process or food/drugs/chemicals. These edits to industrial structures are a significant improvement because they provide a more descriptive classification.

One modification change that needs more investigation and has significant potential is when OCC_CLS is changed from residential to unclassified while PRIM_OCC remains the same (Row 8; 3.7% of the total changes). Many of these records seem to be FPs in the USA Structures data. Some represent a misclassification of the ResType model, where the structure is easily recognized as a nonresidential structure but remains difficult to classify. A new domain value for FPs should be added to flag certain geometries for

Table 7. Summary of USGS feedback and edits to USA Structures data over St. John the Baptist Parish, Louisiana.

Changes in OCC_CLS	Changes in PRIM_OCC	Counts of changes	Percentage of total edits
No change	Single-family dwelling to manufactured Home	907	30.00
No change	Manufactured home to single-family dwelling	152	5.03
No change	Unclassified to manufactured home	103	3.41
No change	Light to metals/minerals processing	139	4.60
No change	Light to food/drugs/chemicals	112	3.70
Residential to agriculture	No change	120	3.97
Residential to industrial	Single-family dwelling to food/drugs/chemicals	84	2.78
Residential to unclassified	No change	112	3.70
Unclassified to industrial	Unclassified to food/drugs/chemicals	71	2.35

removal. If all users intended this combination of changed attribution to label FPs, then these insights are very valuable.

Row 3 in Table 7 (residential/unclassified to manufactured home) validates that the ResType model was correct, and demonstrates how analysts can provide more value with a specific classification. Furthermore, the changes highlighted in Row 7 show that some of ORNL’s source data are inaccurate because the residential/single-family dwelling classification comes from either parcel data or census housing data. These types of modifications can help ORNL discover inaccuracies in the source data. Finally, the modifications to occupancy types captured in Row 9 (unclassified to industrial) are an example of validation for the ResType model. When the ResType model produces a “nonresidential” inference, it is labeled as unclassified for both OCC_CLS and PRIM _OCC. These unclassified structures are assumed to be nonresidential; however, ORNL lacks source information to provide a more specific classification. These modifications are invaluable because they validate the model, improve USA Structures, and demonstrate the value of the human-in-the-loop strategy by receiving feedback from USGS.

Most of the feedback on occupancy types was incorporated into these dataset updates. To further enable the automatic feedback ingestion workflow, the structure geometries are being updated in the database. The received feedback will be further characterized by the morphology features derived from the updated geometries, allowing new feedback on occupancy types to be automatically verified via the associated morphological features.

6. ONGOING WORK

Our team has made significant strides in updating the USA Structures dataset, leveraging cutting-edge technologies and methodologies. We have successfully pioneered Automatic feature extraction (AFE) implementation using state-of-the-art foundation models, modernized our postprocessing workflow, and integrated population estimates into this dataset refresh. These achievements not only enhance the current data quality but also pave the way for future innovations.

Looking ahead, we have identified several promising research directions that demand in-depth investigation and novel approaches to improve scalability and performance. The following three ongoing initiatives are expected to significantly contribute to the next phase of USA Structures data updates, further solidifying our commitment to continuous improvement and technological advancement.

6.1 CHANGE DETECTION WITH SYNTHETIC APERTURE RADAR IMAGES

AFE for detecting buildings requires the use of high-resolution satellite imagery. Orchestrating and monitoring the AFE workflows for all of the very high-resolution optical imagery over the entire US toward updating the USA Structures dataset is both compute- and data-intensive. This research effort is geared toward alleviating the compute- and data-intensive nature of this process by leveraging a tipping and cuing mechanism with Sentinel 1 synthetic aperture radar (SAR) imagery. Sentinel-1 SAR is a satellite mission by the European Space Agency that provides radar imaging for land and ocean monitoring. Owing to its all weather day–night sensing capability, SAR is being used for a wide spectrum of Earth observation applications, including change monitoring. Considering factors such as abundant availability of time-series Sentinel 1 SAR imagery, ease of access, and its coarser resolution (ground-sampling distance of 10 m), this modality is leveraged to detect changes throughout the United States. The central idea of this approach is to inform the AFE workflow to request and process very high-resolution optical imagery only over the regions with detected changes, toward updating the USA Structures dataset.

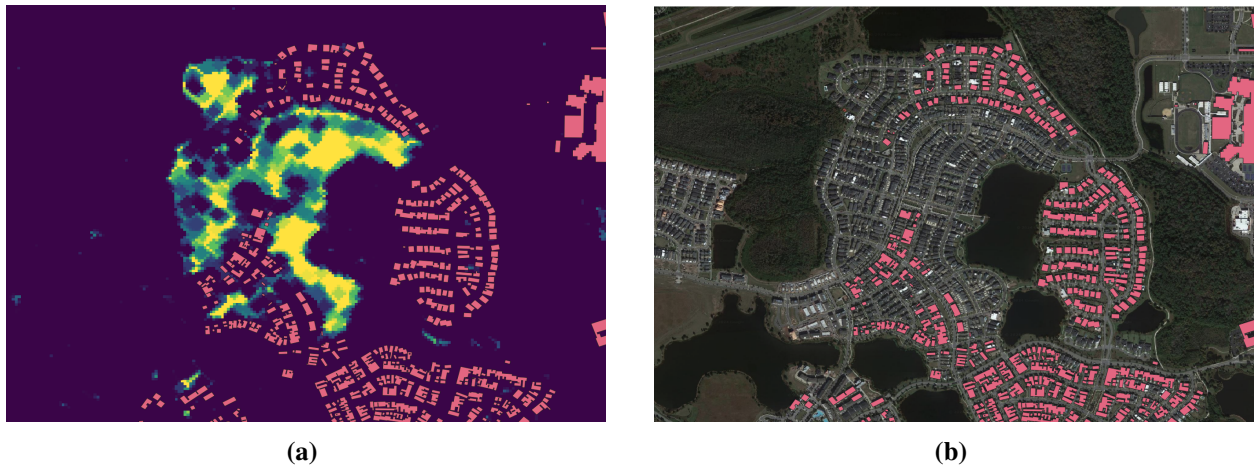


Figure 10. Preliminary results of change detection workflow. (a) Green-yellow blobs represent changes in a region in Orlando, Florida. (b) Google Earth image of the same region. Pink polygons represent USA Structures building footprints before the update.

6.1.1 Methodology

In another report (Ballinger 2024), the authors put forth a pixel-wise t-test based algorithm for persistent building damage detection using backscatter amplitude with pre- and post-event Sentinel 1 SAR imagery. A t-test is a statistical test used to determine whether a significant difference exists between the arithmetic means of two groups, which may be related in certain features. This approach was used to detect changes on the ground between two temporal periods: (1) the last 2–4 months and (2) 2–4 months from the acquisition date of very high-resolution imagery that was used for the previous USA Structures AFE workflow.

For prototyping, we use Google Earth Engine’s data catalog to access time-series imagery of Sentinel 1 SAR ground-resolution distance product. Cloud-optimized aggregation of time-series imagery provided by Google Earth Engine (Amani et al. 2020) was used to compute the pixel-wise t-test for the two sets of imagery toward detecting changes. A postprocessing workflow that performs vectorization and filtering at scale was implemented to generate the vector change maps. Ancillary data sources such as National Land Cover Database (Wickham et al. 2023) by USGS were used to filter out non-built-up related changes.

6.1.2 Preliminary Results

Figure 10 depicts preliminary results of the implemented change-detection workflow. The figure compares the detected changes with building footprints from the previous USA Structures dataset (depicted by pink polygons) over Orlando, Florida. The green-yellow blobs represent detected changes over a newly constructed building complex.

Limitations: Although the preliminary results look promising, some limitations were observed. The current approach of change detection is not restricted to detecting changes over built-up areas—it also flags changes over other land use/land cover classes such as vegetation and bodies of water. Figure 11 depicts instances when the change-detection workflow classified a body of water and vegetation as changes. Multiple instances of changes were detected in regions with a large expanse of vegetation, such as state parks and nature reserves.

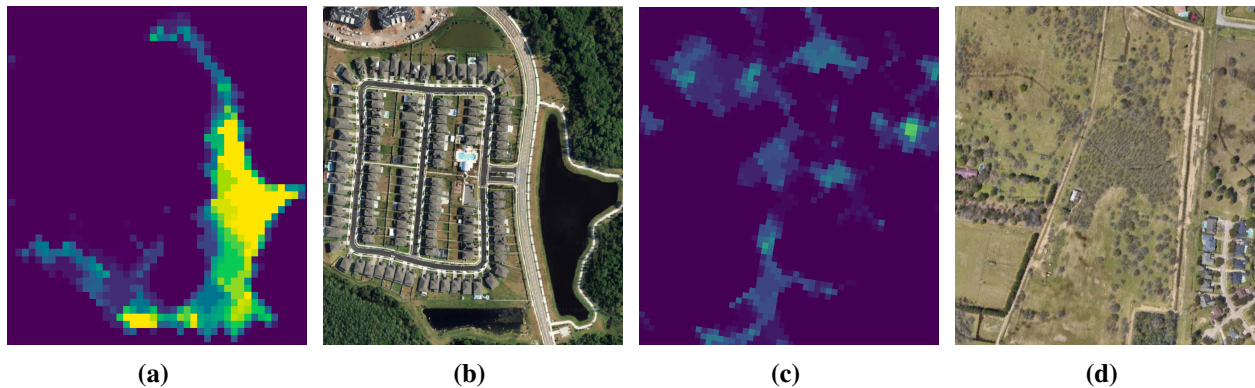


Figure 11. Limitations of change detection workflow. (a) Green-yellow blobs represent change detected in a region in Orlando, Florida. (b) Google Earth image of the region shows the detected change is a body of water. (c) Green-yellow blobs represent change detected in a region in Houston, Texas. (d) Google Earth image over the region shows that the detected change is vegetation.

6.1.3 Future Directions

The change-detection workflow will be improved by incorporating other data sources such as Global Surface Water from the European Commission’s Joint Research Center (Pekel et al. 2016) and higher-resolution

land cover products to eliminate non-built-up related changes. Future work will also involve exploring the optical modality–Sentinel 2 imagery toward improving change-detection performance. The change-detection performance will be quantified to further assess and improve the current workflow. To eliminate the dependency on Google Earth Engine, optimized aggregation of time-series Sentinel 1 imagery will be implemented for computing the pixel-wise t-test for change detection.

6.2 RESTYPE MODEL TRAINING

The ResType model has been used to create previous versions of the USA Structures dataset to address gaps in parcel-derived building purpose information conflated to building footprints. This model imputes missing information as either residential or nonresidential using building morphologies and a gradient-boosting trees classification system. Previously, training datasets for the ResType model were created and prepared under the assumption that equal representation of residential and nonresidential structures would yield optimal classification performance. This update cycle involved investigating the effects of class imbalance between residential and nonresidential classes on classification efficacy. Methods for managing this imbalance were also explored to guide future enhancements of the ResType model. The approach included a multivariate test to determine the most effective method for managing class imbalance, comparing random oversampling, random undersampling, Synthetic Minority Oversampling Technique (SMOTE), Borderline-SMOTE, and Adaptive Synthetic Sampling Approach (ADASYN). Additionally, a sensitivity analysis was conducted to assess the effect of varying class imbalance proportions on the classification performance in training datasets. The study was conducted both at a regional level, where all nine states made up the dataset, and at a state-by-state level.

6.2.1 Methods and Experimental Setup

Data: The study used the USA Structures Phase 2 dataset, specifically data from Texas, Louisiana, Mississippi, Alabama, Georgia, Florida, South Carolina, North Carolina, and Virginia. The labels were the same as the labels used in the USA Structures 2023 ResType model, in which 25 parcel land-use codes were aggregated to residential and 160 land-use codes were used to create the nonresidential class. This strategy resulted in a label dataset of 36,586,077 units: 31,052,538 (85%) residential and 5,533,539 (15%) nonresidential.

Classification Model: The primary classification model employed in the study was a random forest classifier (Breiman 2001). The maximum depth of each tree constructed in the forest was set to 12, allowing each tree to grow large enough to capture the complexities of the morphology feature set without becoming overly deep to the point at which overfitting may become a concern. All other parameters were left to their default settings as set by the sklearn library (Pedregosa et al. 2011).

Classification Validation: The dataset was initially prepared for modeling by dividing it into two subsets, training and testing, using a 70/30 split. The feature scales of the training data were standardized by calculating scaling parameters on the training set and applying the same parameters to the testing set. After standardization, the testing data were left unchanged to evaluate the performance of different sampling methods when tested on the original class distribution.

Estimating Near-Optimal Class Imbalance and Sampling Techniques: To determine the near-optimal class imbalance proportion for training the ResType model, a Bayesian optimization hyperparameter search was conducted for each data sampling technique used to manage class imbalance. The sampling techniques used to sample the training dataset were random oversampling, random undersampling, SMOTE (Chawla et al. 2002), Borderline-SMOTE (Han, Wang, and Mao 2005), and ADASYN (He et al. 2008). These techniques offer a comprehensive array of approaches to handling class imbalance in this dataset, ranging from simple random methods to more sophisticated synthetic data-generation strategies.

Each of these sampling techniques involves certain parameters, as implemented in the imbalanced-learn library (Lemaître, Nogueira, and Aridas 2017), that will be estimated using Bayesian optimization within the following search spaces:

- **SMOTE:** Parameters include *k_neighbors* (the number of nearest neighbors within the minority class for creating synthetic samples) and *sampling_strategy* (the target proportion of the minority class).
 - *k_neighbors*: Discretely uniform from 1 to 16.
 - *sampling_strategy*: Continuously uniform from 0.2 to 1.0.
- **Borderline-SMOTE:** Parameters include *m_neighbors* to identify additional neighbors close to the class boundary, along with *k_neighbors* and *sampling_strategy*.
 - *k_neighbors*: Discretely uniform from 1 to 16.
 - *m_neighbors*: Discretely uniform from 1 to 16.
 - *sampling_strategy*: Continuously uniform from 0.2 to 1.0.
- **ADASYN:** Parameters include *n_neighbors* to focus on areas where the minority class is underrepresented, in conjunction with *sampling_strategy*.
 - *n_neighbors*: Discretely uniform chosen from 1 to 16.
 - *sampling_strategy*: Continuously uniform from 0.2 to 1.0.
- **Random Oversampler and Undersampler:** Parameters include only *sampling_strategy* for both techniques, increasing the minority class in oversampling or decreasing the majority class in under-sampling.
 - *sampling_strategy*: Continuously uniform from 0.2 to 1.0.

Multivariate Test of Class Imbalance Management Techniques: A multivariate test employing Monte Carlo simulations was conducted to determine the most effective sampling technique for building use-type classification. The multivariate test was designed to evaluate the average performance of a building use-type classifier when trained using different sampling techniques to balance the training dataset. To set up the multivariate test, a series of Monte Carlo simulations were conducted for the random oversampling, random undersampling, SMOTE, Borderline-SMOTE, and ADASYN sampling techniques as treatments to the class imbalance observed in the original labeled feature dataset. Each set of simulations used near-optimal parameters, estimated from Bayesian optimization on a 10% subsample of the dataset. The effectiveness of these sampling techniques in addressing class imbalance was assessed by comparing the F1 score performance of each classifier across all simulated runs. This comparison included a control model where no sampling technique was applied.

Sensitivity Testing: The ability of the ResType model to classify buildings when exposed to varying amounts of class imbalance proportions within the training data was tested using the best-performing sampling technique identified from the multivariate testing of class-imbalance management techniques.

The sensitivity testing focused on evaluating the effects of varying class imbalance proportions, specifically the “*sampling_strategy*” parameter, on building use-type classification. This testing was structured similarly to the previously described multivariate tests but focused on assessing the performance across a range of class imbalance sizes. For the regional scale, proportions ranged from 0.2 to 1.0, increasing in increments of 0.1, from highly imbalanced to perfectly balanced datasets. However, at the state level, the range commenced from 0.3 because several states exhibited naturally higher class imbalances than 0.2. For

each class imbalance proportion, a Monte Carlo simulation was conducted. These simulations were designed to mirror the setup of the multivariate tests, employing the same classifier configuration and data preparation steps. The primary goal was to observe how variations in the class-imbalance proportion affected the classifier’s performance, particularly in terms of the F1 score. The outcome of the Monte Carlo simulations for each tested “sampling_strategy” value was the average F1 score, aggregating the results of all runs.

6.2.2 Results and Discussion

Table 8 lists the results from the Bayesian optimization for each of the sampling techniques used to manage class imbalance in the ResType training dataset. For the region-wide study area, near-optimal class imbalance is achieved using a 2:1 ratio; that is, the ResType model performs best when the training dataset contains two residential instances for each nonresidential instance. Figure 12 shows that, across the region-wide study area, simple random oversampling and random undersampling techniques tend to be the best class-imbalance management technique for preparing the training dataset for the ResType model.

The results for the state-by-state study areas, listed in Table 9, suggest that simple random oversampling and random undersampling techniques tend to outperform the more sophisticated methods tested. Moreover, the state-by-state results in Figure 13 show that introducing some level of class imbalance into the training dataset enhances the ResType model’s classification performance. However, the state-by-state results indicate that the near-optimal class imbalance proportions vary widely for each state. Therefore, the observed 2:1 ratio identified in the regional study is not a heuristic that can generally be used for residential or nonresidential classification. Instead, this near-optimal imbalance must be determined based on more localized geographies.

These results have implications for the next version of the ResType model. First, the results challenge the previous assumption that equal representation of residential and nonresidential classes in the training dataset yields better classification performance. Figures 12 and 13 show that equal representation of these two classes in the training dataset often results in significantly worse classification performance compared with when imbalance has been introduced into the training dataset. Second, the observed differences in near-optimal class imbalance state-by-state in Figure 13 suggest that the ResType model may yield better performance if subregional-level models were created, rather than utilizing a single generalized model for extensive geographic regions.

Table 8. Bayesian optimization results for various sampling techniques across regions. This table includes parameters optimized for the study area and the states within the study area (*k_n* represents *k_neighbors*, *m_n* denotes *m_neighbors*, *n_n* signifies *n_neighbors*, and *proportion* refers to the *sampling_strategy*).

Region	SMOTE		Borderline-SMOTE			ADASYN		Random oversampler	Random undersampler
	<i>k_n</i>	<i>proportion</i>	<i>k_n</i>	<i>m_n</i>	<i>proportion</i>	<i>n_n</i>	<i>proportion</i>	<i>proportion</i>	<i>proportion</i>
Study area	6	0.5496	10	4	0.4423	5	0.4714	0.5303	0.5005
Mississippi	5	0.7815	10	8	0.6464	9	0.7481	0.9052	0.7964
Virginia	7	0.5733	12	9	0.4299	10	0.5157	0.5764	0.5055
South Carolina	2	0.7278	2	13	0.5638	6	0.6622	0.7268	0.5924
Alabama	4	0.8099	10	2	0.7369	1	0.6217	0.7380	0.6855
Florida	1	0.4761	3	12	0.3663	16	0.4050	0.4797	0.3979
Georgia	9	0.4299	10	7	0.3383	4	0.3578	0.4937	0.4100
Louisiana	7	0.6874	4	14	0.5182	13	0.5638	0.7356	0.6127
North Carolina	2	0.5556	5	8	0.4505	13	0.4859	0.6127	0.5182
Texas	2	0.5009	11	9	0.3778	9	0.3979	0.4704	0.4715

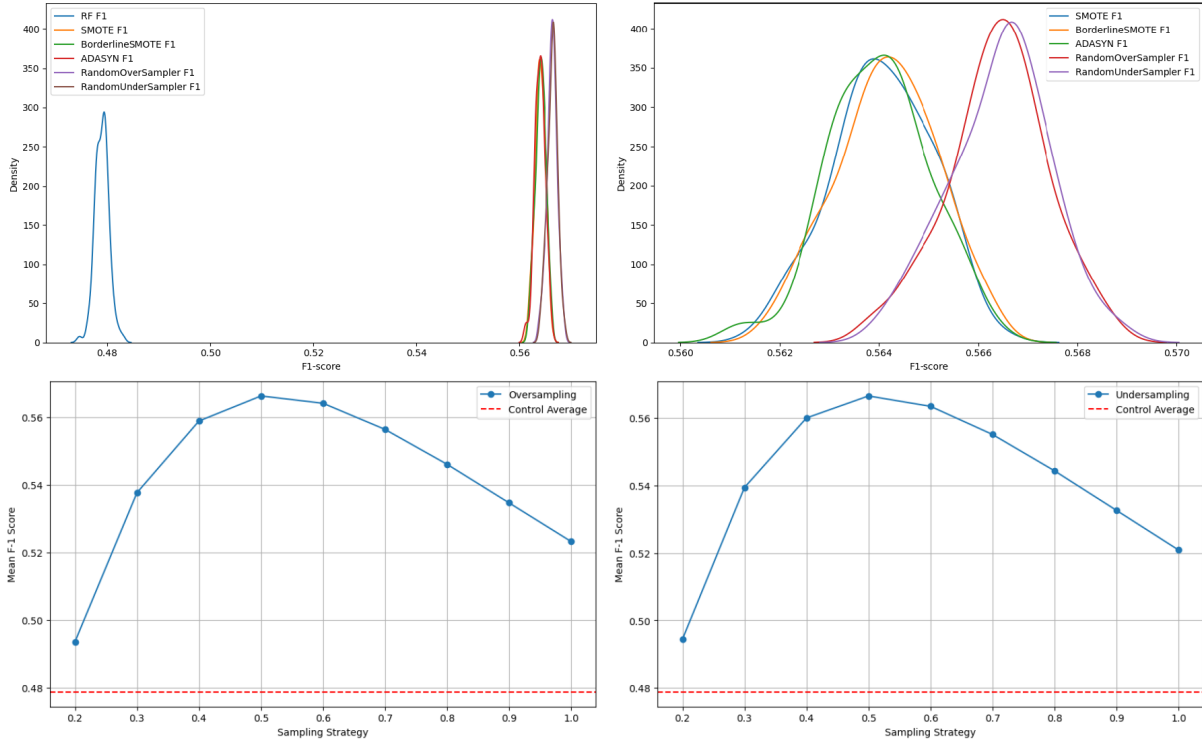


Figure 12. Multivariate and sensitivity testing results for the study area. (top left) KDE plots of the Monte Carlo simulations for the sampling techniques tested and the control model. (top right) The sampling techniques tested. (bottom) Results of the sensitivity analysis: different sampling proportions are used with the random over- and undersampling techniques.

6.3 IMPROVEMENT TO AUTOMATIC QA/QC

Because of the significant improvements to the building feature extraction, the verification and validation model (VVM) will need to be retrained. The retraining of the VVM will require collecting FP examples. This effort is one reason for the parallelization of the workflow to allow more time for manual review. The FP mask can be used to label geometries as FPs. Once enough samples are collected, features from the Geographic Augmentation of Extracted Building Features Tool (GAUNTLET) will be generated from the updated geometries, allowing the VVM to assess the new model architecture output and using the latest version of GAUNTLET morphologies to do so.

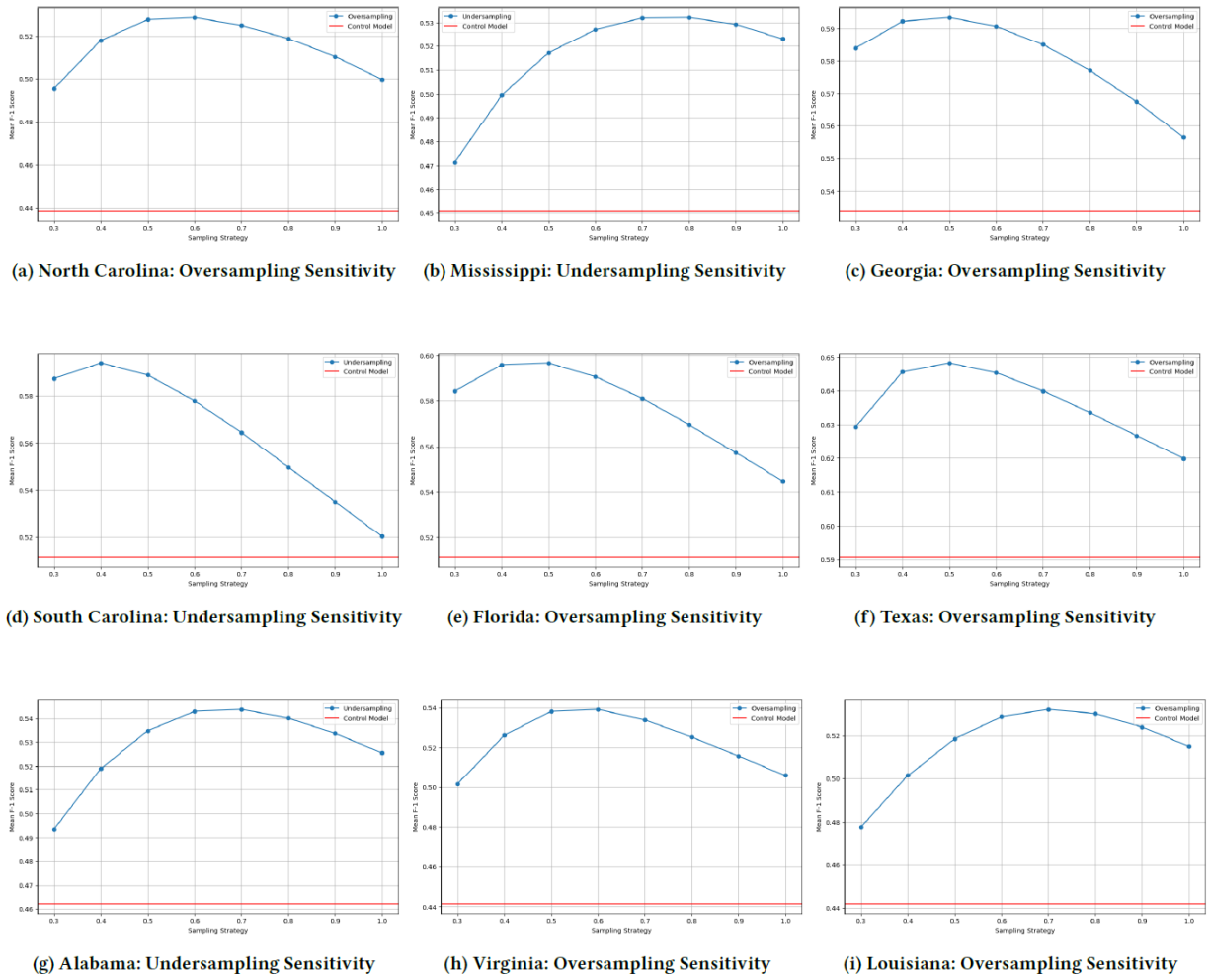


Figure 13. State-by-state sensitivity testing results. For each state, the top-performing sampling method’s sensitivity is shown.

Table 9. Summary of state-by-state multivariate tests. Shown are the average F1 Scores from each multivariate test using the near-optimal parameters derived from Bayesian optimization.

State	Control	SMOTE	B-SMOTE	ADASYN	Random oversampler	Random undersampler
Texas	0.5903	0.6473	0.6463	0.6473	0.6485	0.6485
Louisiana	0.4422	0.5267	0.5279	0.5259	0.5316	0.5314
Alabama	0.4614	0.5388	0.5393	0.5367	0.5427	0.5431
Mississippi	0.4513	0.5268	0.5267	0.5265	0.5328	0.5331
Florida	0.5111	0.5948	0.5941	0.5921	0.5969	0.5938
Georgia	0.5339	0.5918	0.5904	0.5916	0.5940	0.5917
South Carolina	0.4567	0.5362	0.5364	0.5312	0.5394	0.5397
North Carolina	0.4376	0.5267	0.5270	0.5258	0.5283	0.5281
Virginia	0.4425	0.5376	0.5368	0.5374	0.5401	0.5392

7. REFERENCES

- Amani, Meisam, Arsalan Ghorbanian, Seyed Ali Ahmadi, Mohammad Kakooei, Armin Moghimi, S Mohammad Mirmazloumi, Sayyed Hamed Alizadeh Moghaddam, Sahel Mahdavi, Masoud Ghahremanloo, Saeid Parsian, et al. 2020. "Google earth engine cloud computing platform for remote sensing big data applications: A comprehensive review." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13:5326–5350.
- Ballinger, Ollie. 2024. "Open Access Battle Damage Detection via Pixel-Wise T-Test on Sentinel-1 Imagery." *arXiv preprint arXiv:2405.06323*.
- Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45:5–32. <https://doi.org/10.1023/A:1010933404324>. <https://doi.org/10.1023/A:1010933404324>.
- Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. "SMOTE: Synthetic Minority Over-sampling Technique." *Journal of Artificial Intelligence Research* 16 (June): 321–357. ISSN: 1076-9757. <https://doi.org/10.1613/jair.953>. <http://dx.doi.org/10.1613/jair.953>.
- Dias, Philipe, Aristeidis Tsaris, Jordan Bowman, Abhishek Potnis, Jacob Arndt, H. Lexie Yang, and Dalton Lunga. 2024. "OREole-FM: billion-parameter foundation models for high-resolution satellite imagery applications." *32nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2024)*, <https://doi.org/10.1145/3678717.3691292>. <https://doi.org/10.1145/3678717.3691292>.
- FEMA, Building Science Division. n.d. "Building Code Adoption Portal." Accessed: 2024-07-30.
- Getis, Arthur, and J. K. Ord. 1992. "The Analysis of Spatial Association by Use of Distance Statistics." *Eprint*: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1538-4632.1992.tb00261.x>, *Geographical Analysis* 24 (3): 189–206. <https://doi.org/https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1538-4632.1992.tb00261.x>.
- Han, Hui, Wenyuan Wang, and Binghuan Mao. 2005. "Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning." In *International Conference on Intelligent Computing*. <https://api.semanticscholar.org/CorpusID:12126950>.
- Hauser, Taylor, Jessica Moehl, Erik Schmidt, Daniel Adams, Matthew Whitehead, Bennett Morris, and Lexie Yang. 2023. *USA Structures Phase 2 Technical Report*. ORNL Report. Oak Ridge National Laboratory, August.
- He, Haibo, Yang Bai, Eduardo A. Garcia, and Shutao Li. 2008. "ADASYN: Adaptive synthetic sampling approach for imbalanced learning." In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 1322–1328. <https://doi.org/10.1109/IJCNN.2008.4633969>.
- Lemaître, Guillaume, Fernando Nogueira, and Christos K. Aridas. 2017. "Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning." *Journal of Machine Learning Research* 18 (17): 1–5. <http://jmlr.org/papers/v18/16-365.html>.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12:2825–2830.

- Pekel, Jean-François, Andrew Cottam, Noel Gorelick, and Alan S Belward. 2016. “High-resolution mapping of global surface water and its long-term changes.” *Nature* 540 (7633): 418–422.
- Stewart, Robert, Marie Urban, Samantha Duchscherer, Jason Kaufman, April Morton, Gautam Thakur, Jesse Piburn, and Jessica Moehl. 2016. “A Bayesian machine learning model for estimating building occupancy from open source data.” *Natural Hazards* 81:1929–1956.
- Stipek, Clinton, Taylor Hauser, Daniel Adams, Justin Epting, Christa Brelsford, Jessica Moehl, Philippe Dias, Jesse Piburn, and Robert Stewart. 2024. “Inferring building height from footprint morphology data.” *Scientific Reports* 14 (1): 18651.
- Swan, Benjamin, Joe Pyle, Darrell Roddey, Lexie Yang, Amy Rose, and Melanie Laverdiere. 2024. “ORBITaL-Net: A labeled training library for large-scale building feature extraction.” [submitted for publication], *Scientific Data*.
- Tuccillo, Joseph, Robert Stewart, Amy Rose, Nathan Trombley, Jessica Moehl, Nicholas Nagle, and Budhendra Bhaduri. 2023. “UrbanPop: A spatial microsimulation framework for exploring demographic influences on human dynamics.” *Applied Geography* 151:102844.
- Tuccillo, Joseph V. 2024. “Downscaling Synthetic Populations to Realistic Residential Locations.” In *US-RSE 2023 Conference Proceedings*. <https://doi.org/10.5281/zenodo.10420984>. <https://doi.org/10.5281/zenodo.10420984>.
- Urban, Marie, Robert Stewart, Scott Basford, Zachary Palmer, and Jason Kaufman. 2023. “Estimating building occupancy: a machine learning system for day, night, and episodic events.” *Natural Hazards* 116 (2): 2417–2436.
- Wickham, James, Stephen V Stehman, Daniel G Sorenson, Leila Gass, and Jon A Dewitz. 2023. “Thematic accuracy assessment of the NLCD 2019 land cover for the conterminous United States.” *GI-Science & Remote Sensing* 60 (1): 2181143.
- Yang, Hsiuhan Lexie, Melanie Laverdiere, Taylor Hauser, Benjamin Swan, Erik Schmidt, Jessica Moehl, Andrew Reith, et al. 2024. “A baseline structure inventory with critical attribution for the US and its territories.” *Scientific Data* 11, no. 1 (May): 502. ISSN: 2052-4463. <https://doi.org/10.1038/s41597-024-03219-x>. <https://doi.org/10.1038/s41597-024-03219-x>.

