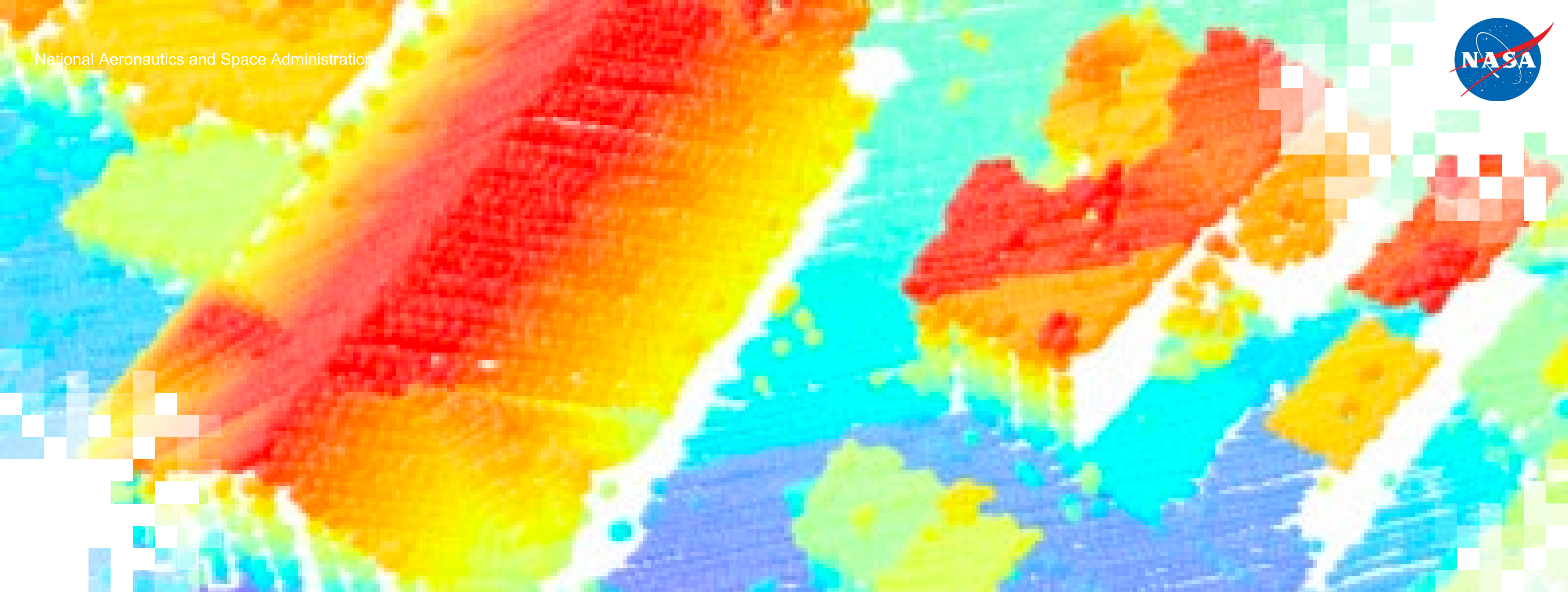
National Aeronautics and Space Administration

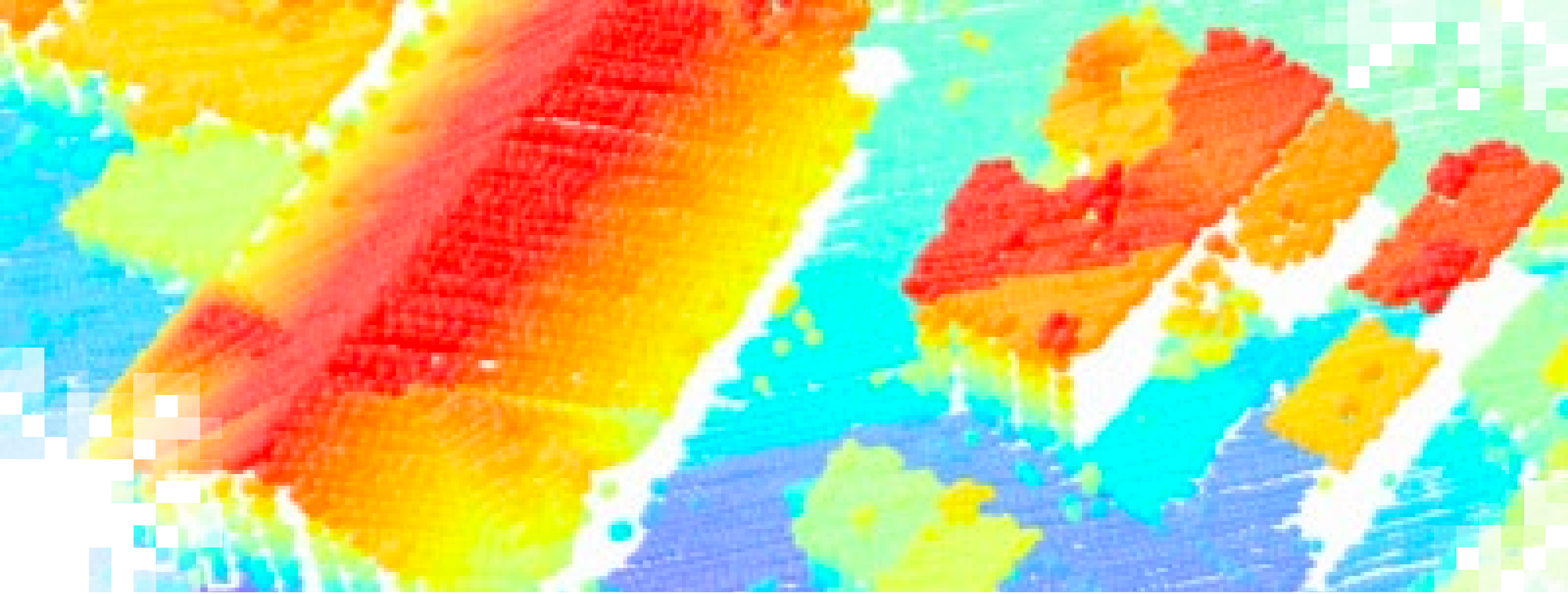


Transforming Earth Observation (EO) Data into Building Infrastructure Data Sets for Disaster Risk Modeling

Part 2: Building Infrastructure Data Sets for Disaster Risk Modeling

Greg Yetman (CIESIN, Columbia University), Juan Martinez (CIESIN, Columbia University) Taylor Hauser (Oak Ridge National Laboratory [ORNL]), Amy Rose (ORNL), Philipe Ambrozio Dias (ORNL)

October 5, 2023



Transforming Earth Observation (EO) Data into Building Infrastructure Data Sets for Disaster Risk Modeling

Overview

Why is Climate Risk Assessment Important?

- Even with drastic reduction in carbon emissions, short and medium-term impacts are inevitable
- Climate change impacts and risks are becoming increasingly complex and more difficult to manage (IPCC AR6, 2022).
- Climate change impacts on human infrastructure are not well understood and vary drastically by location
- Understanding community-specific risks to climate change is critical to evaluating adaptation strategies

"You can't stop the waves, but you can learn to surf." - Jon Kabat-Zinn

Credit: Scott Pena

Training Outline

Part 1	Part 2	Part 3
Development of Regional Exposure Data with EO	**Development of Site-Specific Exposure Data with EO**	Assessing Utility and Communicating Uncertainty
October 03, 2023	**October 05, 2023**	October 10, 2023
!0:00-12:00 EDT (UTC-4)	**!0:00-12:00 EDT (UTC-4)**	!0:00-12:00 EDT (UTC-4)

Homework

Opens October 10 – Due October 24 – Posted on Training Webpage

A certificate of completion will be awarded to those who attend all live sessions and complete the homework assignment(s) before the given due date.

Prerequisites

- Fundamentals of Remote Sensing
- Basics of GIS and Databases
- Basic Statistics and Sampling

How to Ask Questions

- Please put your questions in the Questions box and we will address them at the end of the webinar.
- Feel free to enter your questions as we go. We will try to get to all of the questions during the Q&A session after the webinar.
- The remainder of the questions will be answered in the Q&A document, which will be posted to the training website about a week after the training.

Part 2 – Trainers

Greg Yetman

Associate Director for Geospatial Applications

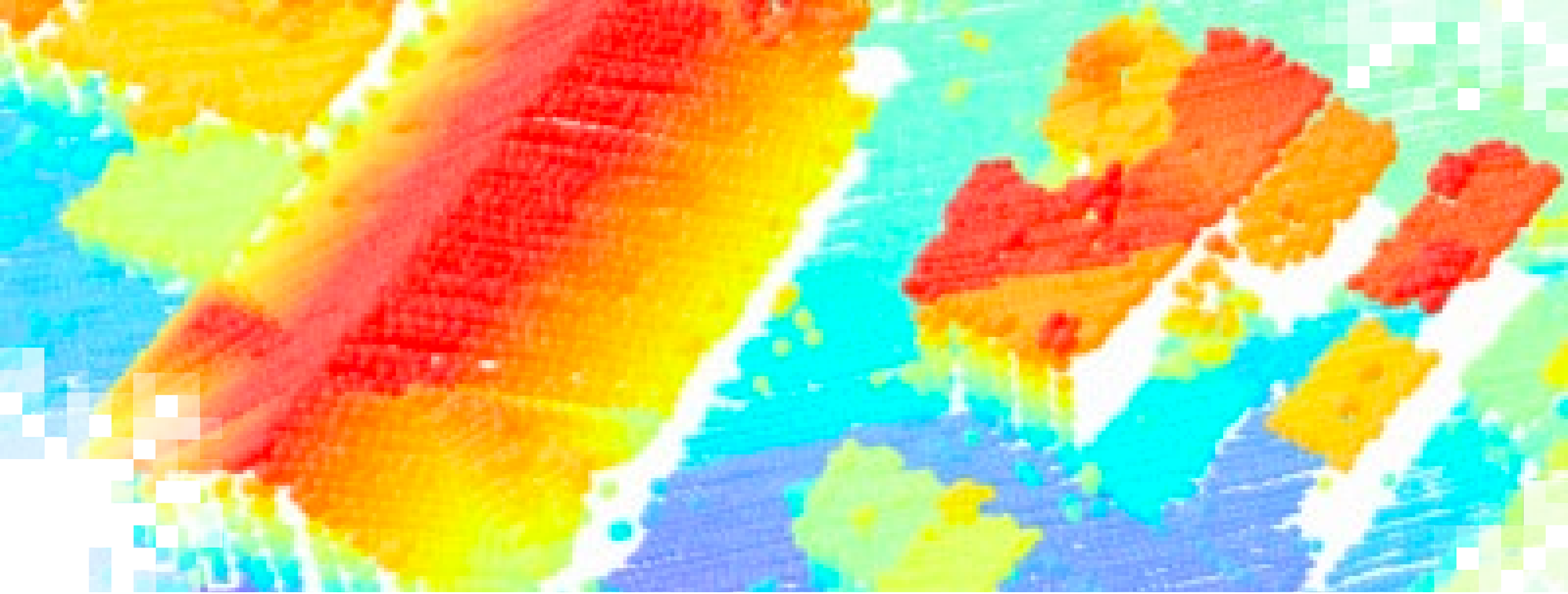
CIESIN, Columbia University

Taylor Hauser

Geospatial Data Analyst

Geospatial Data Modeling Group

Oak Ridge National Laboratory



Transforming Earth Observation (EO) Data into Building Infrastructure Data Sets for Disaster Risk Modeling

Developing a Building-level Exposure Data Set for HAZUS

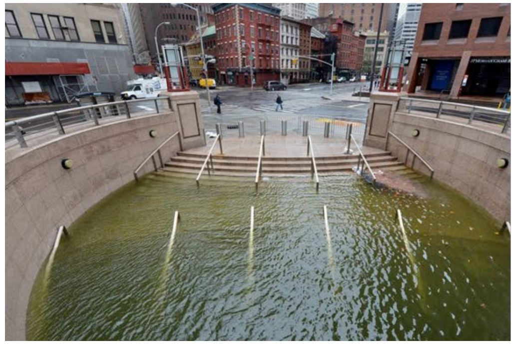
Part 2 Objectives

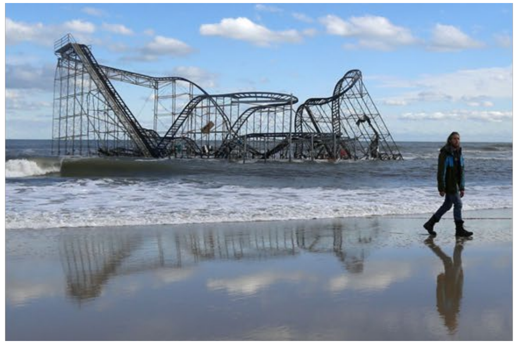
By the end of Part 2, participants will be able to:

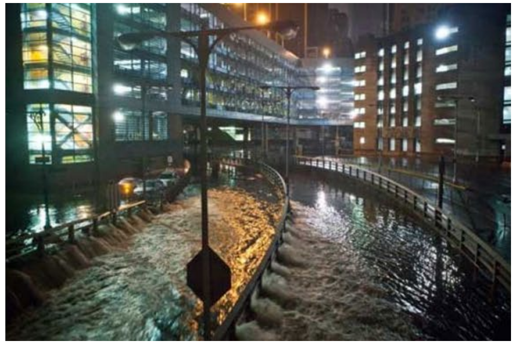
- Understand techniques for attributing infrastructure data
- Identify issues in attribute transfer due to spatial location issues
- Select strategies for transferring attributes
- Identify common sources for infrastructure and building footprint data sources

Developing a Building-level Exposure Data set for HAZUS

- October 2012: New York State and the surrounding area were rocked by Superstorm Sandy.
- Post-storm questions:
  - Were we prepared?
  - What could we do differently?
  - What data are needed for the next storm?
- Risk assessment requires more than the location of population and buildings.

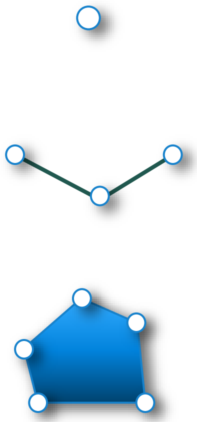






Review of Prior Knowledge

- Geographic Information Systems (GIS) support spatial and attribute query.
- Vector GIS data are comprised of three fundamental types:
  - Points: 0-dimensional objects
  - Lines: 1-dimenstional objects
  - Polygons: 2-dimenstional objects
- Spatial overlay allows different layers to be combined by computing geometric relationships.
  - Which buildings fall inside which tax parcels?



Point: x,y coordinate pair

Line: ordered series of x,y coordinate pairs

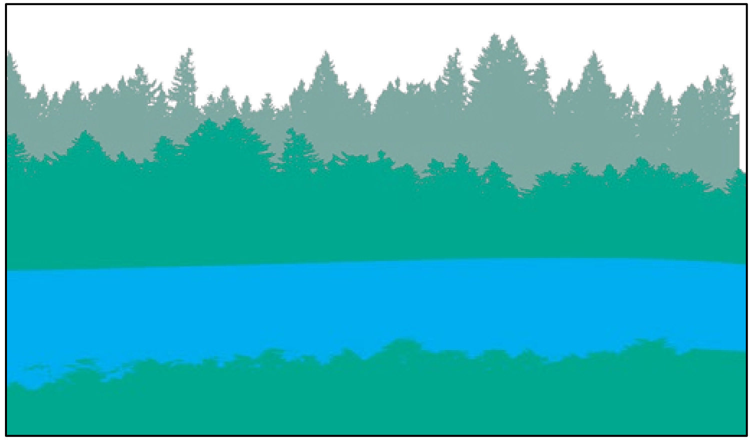
Polygon: closed series of x,y coordinate pairs

CountyName	Rockland
MuniName	Orangetown
Source	Rockland County GIS Division
SourceID	
SourceDate	2007
RoofType	0
InfrType	
OccClass	EDU2

Flood Analysis: HAZUS

Hazards US (HAZUS) can estimate flood impacts in GIS based on data inputs:

- Flood depth
- Building data information
  - Value
  - Occupancy class (residential, industrial, commercial, mixed, institutional)
    - Broken down further, e.g., RES-4
  - Ground-flood elevation (lowest entry point for water)
  - Critical infrastructure



Compiling a State-Wide Building Footprint Database

Integration of building footprints from multiple sources

- Municipal (city/town/county) databases
- Microsoft building footprints
- Footprints extracted from LiDAR data using automated methods
- Manual digitizing from building imagery

Attributing Buildings from Multiple Sources

- Critical infrastructure were compiled from state and federal sources.
  - Critical infrastructure point location data were joined to building footprints by location (nearest polygon)
  - Data were validated through visual inspection and web searches
- Data model altered to handle co-located services in single buildings; e.g., fire and police in the same structure

Critical Infrastructure facilities include:

- College or University
- EMS Station
- Emergency Operation Center
- Fire Station
- Hospital
- Nursing Home
- Place of Worship
- Power Plant
- School
- Wastewater Facility

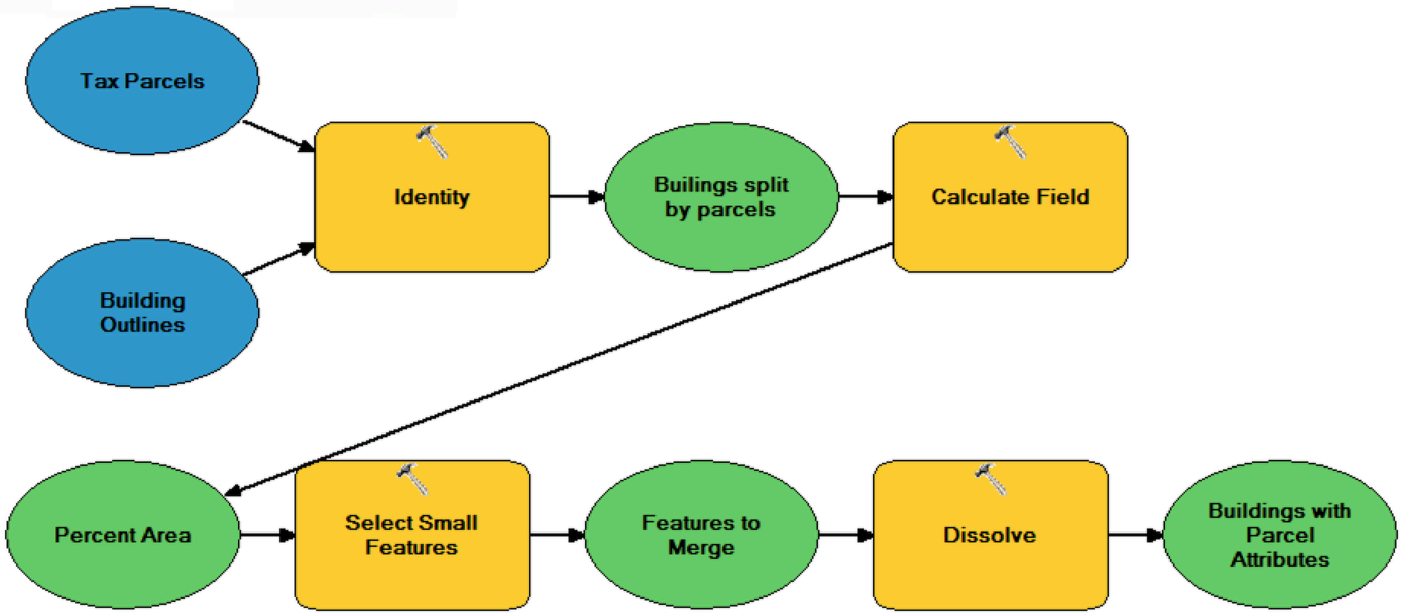
Attributing Buildings from Multiple Sources

- Building assessed value, occupancy class, and year built were obtained from a state-wide tax parcel database.
- Many tax parcels contained multiple buildings.
  - Value divided across buildings by footprint area
- Many buildings crossed tax parcel boundaries.
  - Buildings were split by tax parcels
  - If only a small percentage of area fell on one side of the tax parcel boundary, the small portion was kept with the larger tax boundary

Attributing Buildings from Multiple Sources

Handling buildings that are split across multiple parcels:

1. Split by parcel
2. Calculate percent area by original building ID
3. Select/flag features to merge back with original feature
4. Dissolve selection

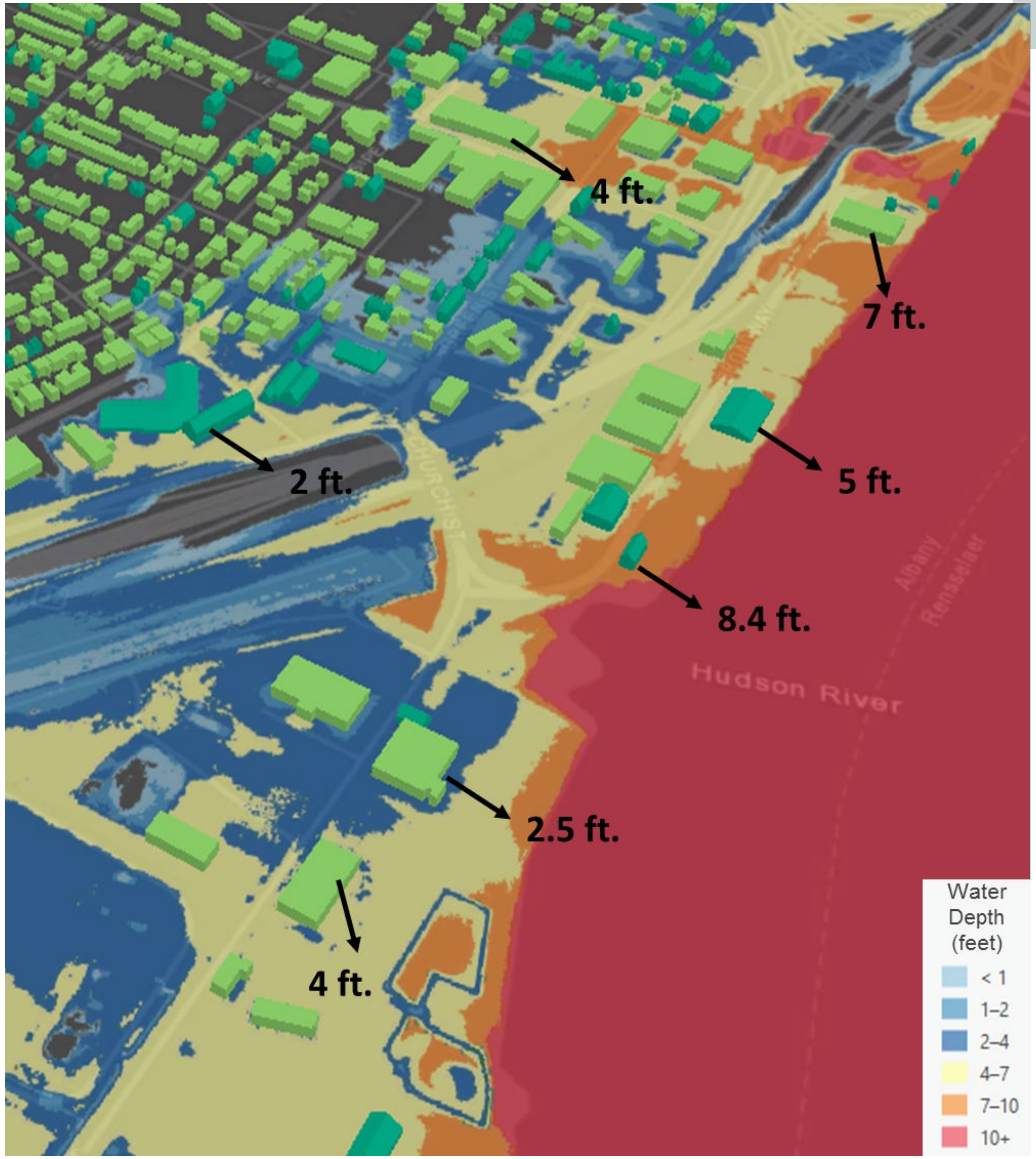


Attributing Buildings from Multiple Sources

- Many parcels have "stacked" polygons: vertical units such as condos or co-ops where each flood has multiple units with separate assessments.
- These were summed (value) and assigned to the building footprint on the lot.
- Mixed use categories had to be created for some buildings.
  - E.g., buildings with ground-floor commercial space and upper-level residences
- Ground floor elevation was generally not available in the building footprint or tax parcel data.
- The "year built" attribute from the tax parcels was used to impute the ground floor elevation.
  - Building codes specify a minimum ground floor; these have changed over the years.
  - The year built was used to lookup the relevant building code minimum elevation.

Impact Assessment

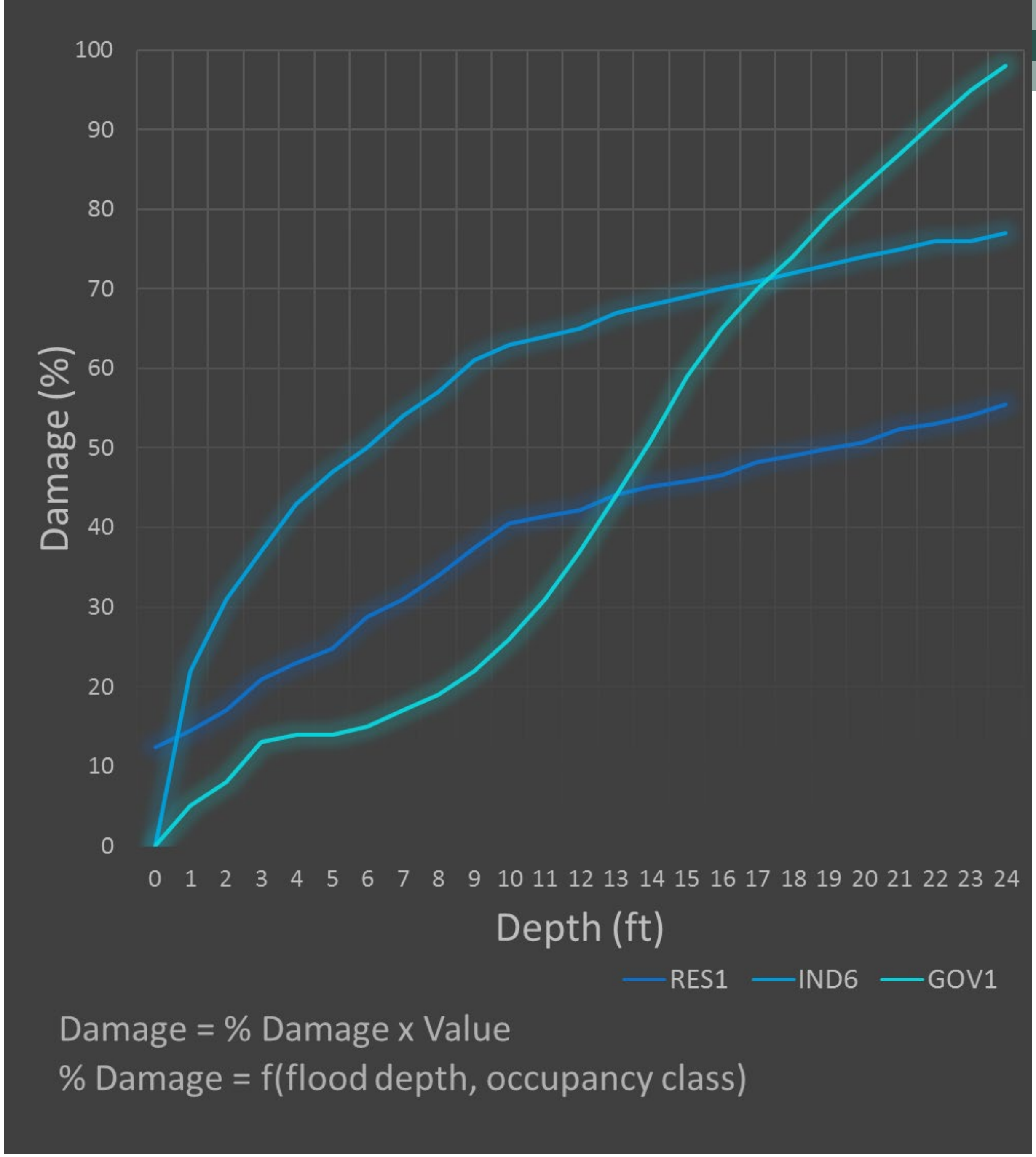
- Attach critical infrastructure and tax parcel information to building footprints.
- Assign flood depth value to each building footprint from flood surface.
  - FEMA DFIRMS for river flooding
  - Stevens Institute modeled flood depths for coastal areas
- Flood depth, occupancy, first floor elevation, and value are key variables.



Impact Assessment

Once all the data is compiled, a comprehensive and detailed flood impact assessment is produced.

- Damage curves based on post-storm impact assessments from across the U.S. (HAZUS)
- Building-level damage estimates (in $USD) were aggregated to municipal units
- Building-level damage category applied per building (slight, moderate, significant, destroyed)



Impact Assessment

- Municipal-level statistics were summed from the building-level analysis.
- Individual buildings were categorized into damage categories for 100 and 500 year floods (1% and 0.2% probability events).
  - None
  - Slight
  - Moderate
  - Substantial

Recommendations for residential building categorized as substantial:

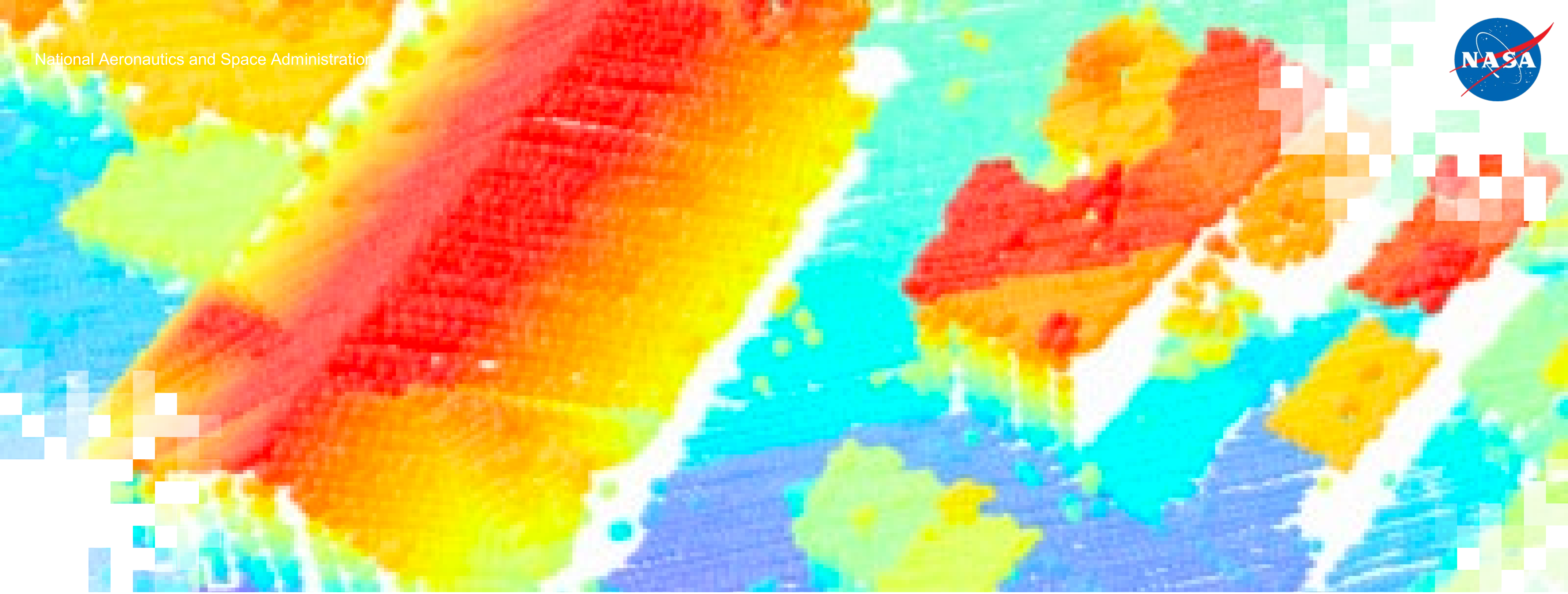
- (1) Obtain flood insurance (2) Elevate furnace, water heater, and electric panel (3) Install "check valves" (4) Incorporate flood openings below BFE, and ensure waterproof building materials.

County	Municipality	Return Period (years)	Building and Contents loss ($)	Number of buildings damaged
Dutchess	Fishkill	100	3,803,235	93
Dutchess	Fishkill	500	9,679,156	367
Dutchess	Pleasant Valley	100	1,590,441	93
Dutchess	Pleasant Valley	500	10,912,804	276

Results

- Building footprints are available for download.
- The flood data and building footprints can be viewed online.

National Aeronautics and Space Administration



Using Earth Observations (EO) to Develop a Building Structures Dataset

Taylor Hauser | Geospatial Data Analyst

Built Environment Characterization Group | Oak Ridge National Laboratory

Outline

- USA Structures Project
  - Background and Motivation
  - Workflow
    - From imagery to polygon
    - From polygon to structure with attribution

Background and Motivation

- In 2016, the U.S. experienced 32 major disasters and six emergency declarations involving floods.
- Emergency preparation, response, and mitigation have been hampered by the lack of accurate, current data on the precise location and elevation of buildings.
- Solution:
  - Build and maintain the nation's (USA) first comprehensive inventory of all structures larger than 450 sq. ft.

USA Structures Project Contributors 2017–Present

- Mark Tuttle
- Melanie Laverdiere
- Lexie Yang
- Taylor Hauser
- Jacob Mckee
- Jessica Moehl
- Bennett Morris
- Joe Pyle
- Ben Swan
- Matthew Whitehead
- Andrew Reith
- Matthew Crockett
- Erik Schmidt
- Daniel Adams
- Darrell Roddy

Progress to Date (September 2017 – August 2023)

Complete: Phase I & II

	Feature Count
Texas	10,466,143
Louisiana	1,859,228
Arkansas	2,489,884
Missouri	1,527,560
Oklahoma	2,323,936
Arizona	2,724,064
New Mexico	986,505
Alabama	2,489,884
Mississippi	1,527,560
Guam	42,663
Hawaii	327,070
Puerto Rico	1,142,054
Georgia	3,757,825
South Carolina	2,286,581
Florida	6,645,067
North Carolina	4,650,575
Illinois	4,639,278
Indiana	3,287,119
Kentucky	2,418,871
Ohio	5,496,516

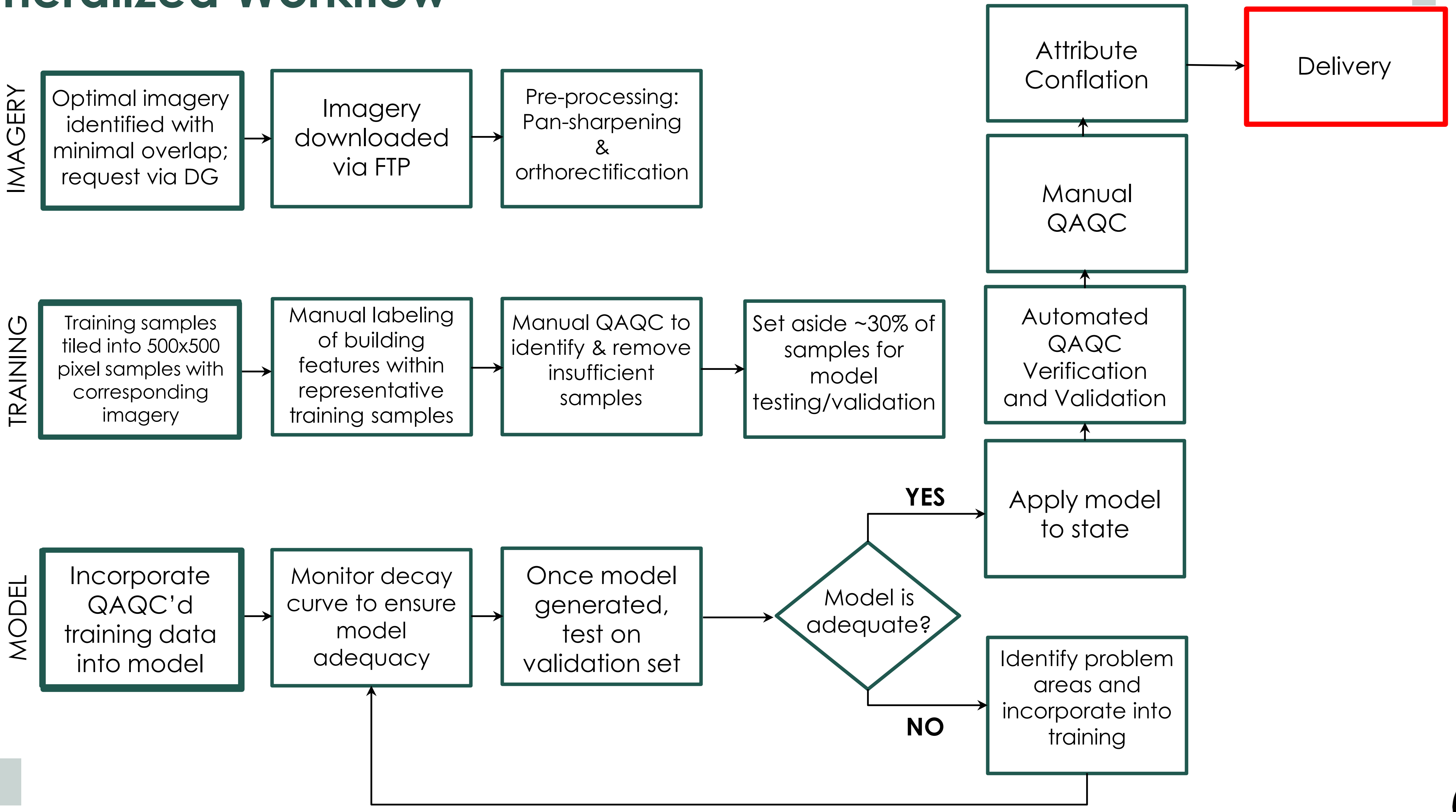
Massachusetts	2,057,472
New Hampshire	558,369
New Jersey	2,467,395
New York	4,847,135
Pennsylvania	4,837,949
Rhode Island	353,194
Vermont	357,733
Virginia	3,124,376
West Virginia	1,072,955
Oregon	1,658,885
Iowa	2,114,520
Michigan	4,782,958
Minnesota	2,801,654
Wisconsin	3,039,604
Virgin Islands	40,726
Kansas	1,600,218
Nebraska	1,178,532
North Dakota	572,242
South Dakota	628,750
Virgin Islands	40,726

Tennessee	3,122,388
California	9,946,076
Nevada	837,251
Idaho	853,335
Washington	2,780,681
Oregon	1,658,885
Connecticut	1,131,222
District of Columbia	58,061
Delaware	371,915
Maine	761,802
Maryland	1,658,164
American Samoa	13,412
Northern Mariana Is.	12,572
Colorado	2,174,948
Montana	767,753
Utah	1,101,597
Wyoming	385,465
Alaska	295,307
Total to date:	**125,583,725**

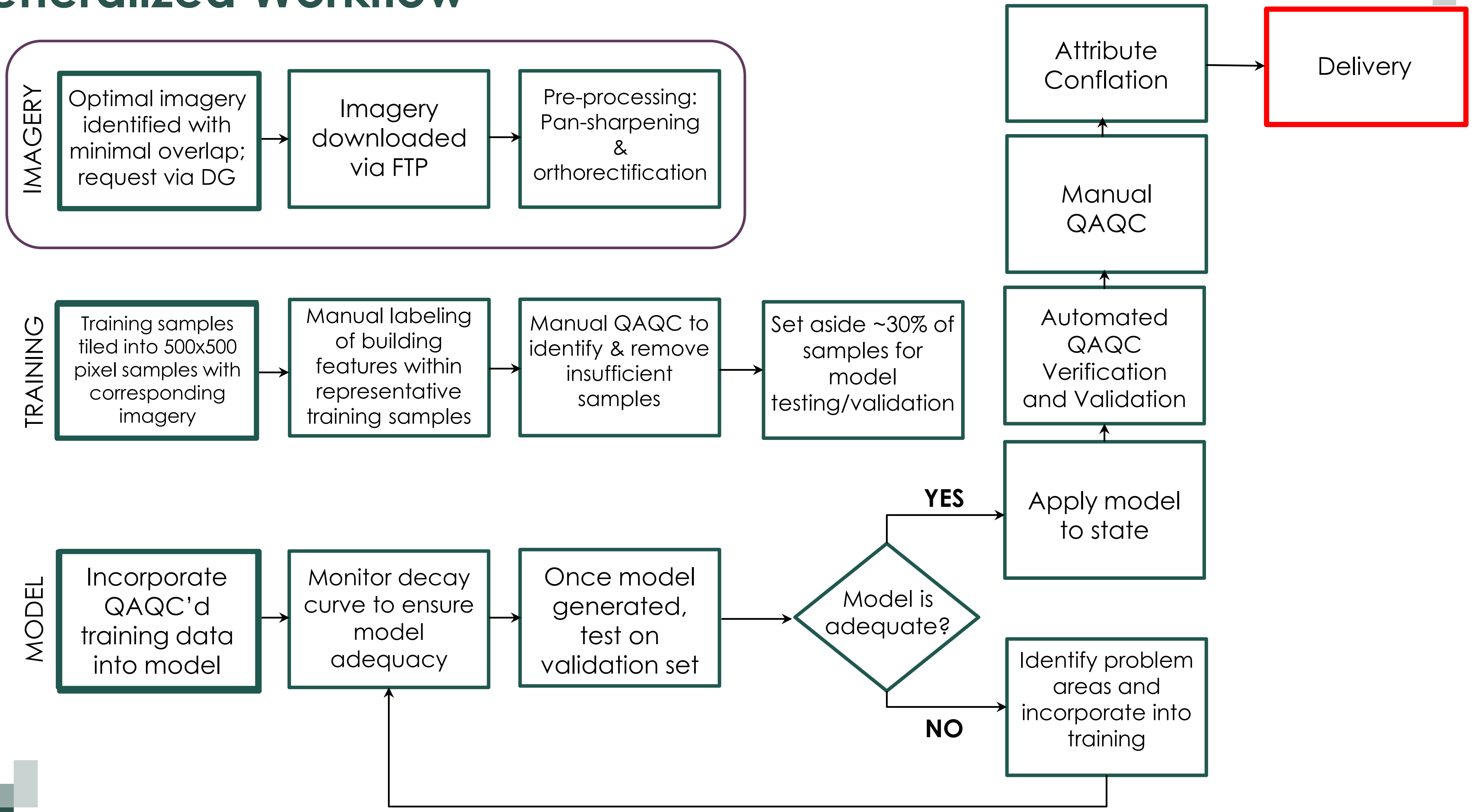
Data Available from FEMA's Disasters Public files on Geoplatform:

https://disasters.geoplatform.gov/publicdata/Partners/ORNL/USA_Structures/

Generalized Workflow

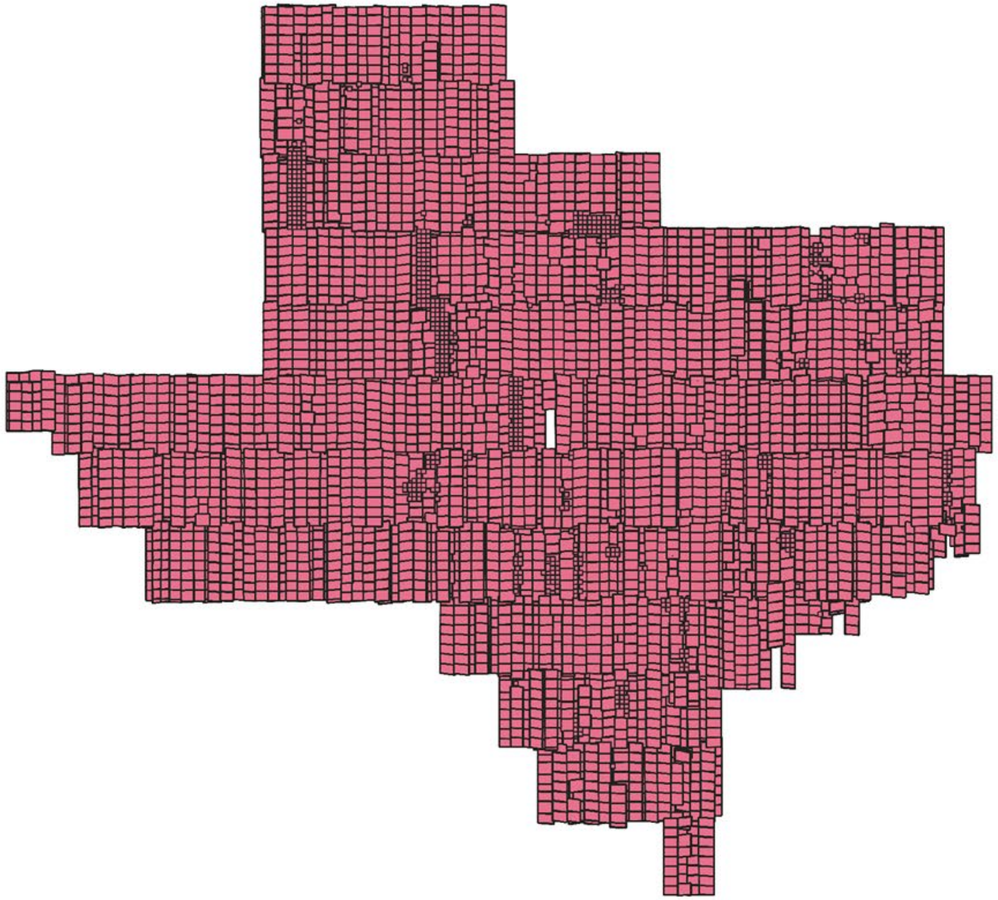


Generalized Workflow

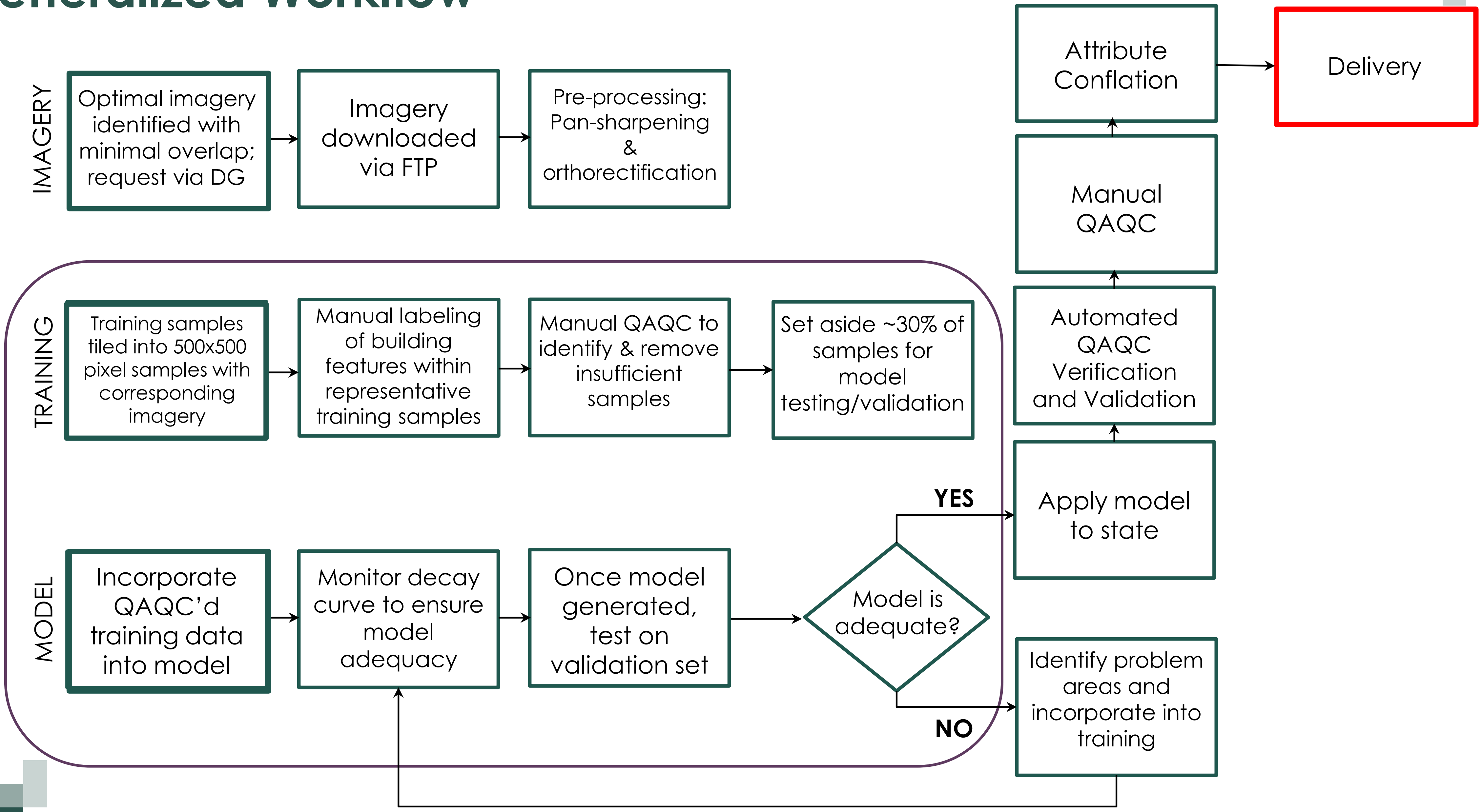


Data Volume Example

- Texas
- 4862 Images
- 3.7 TB
- Pan-sharpened 44 TB



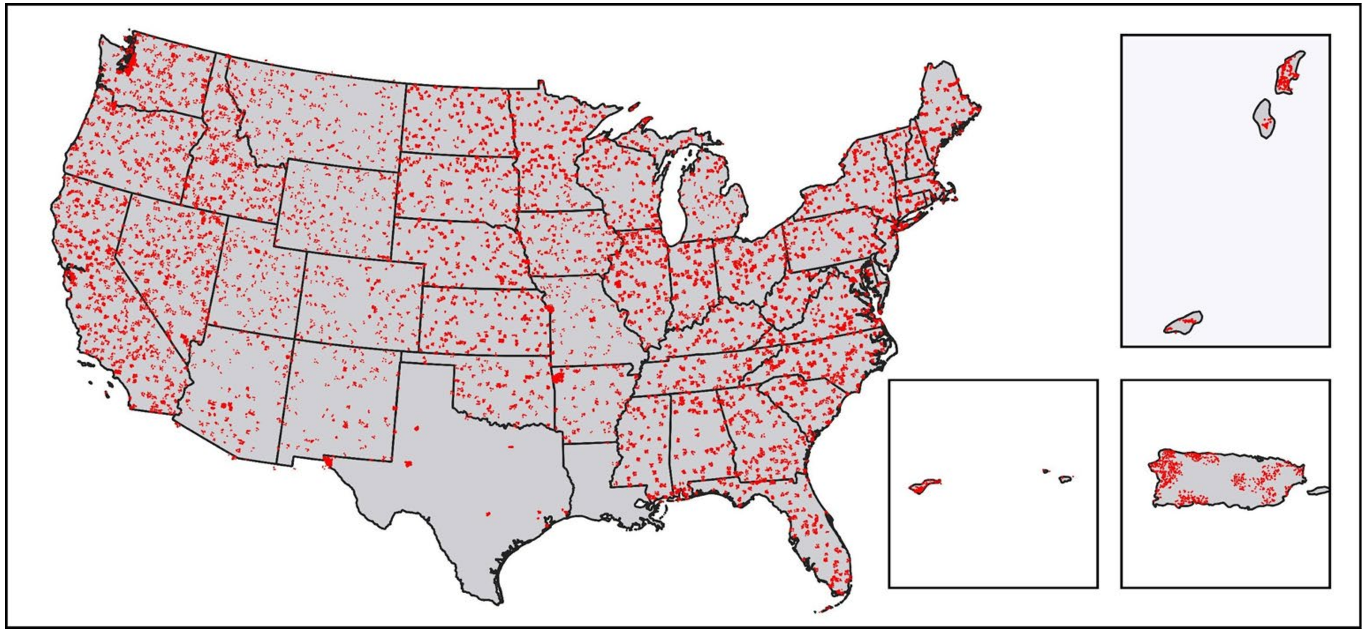
Generalized Workflow

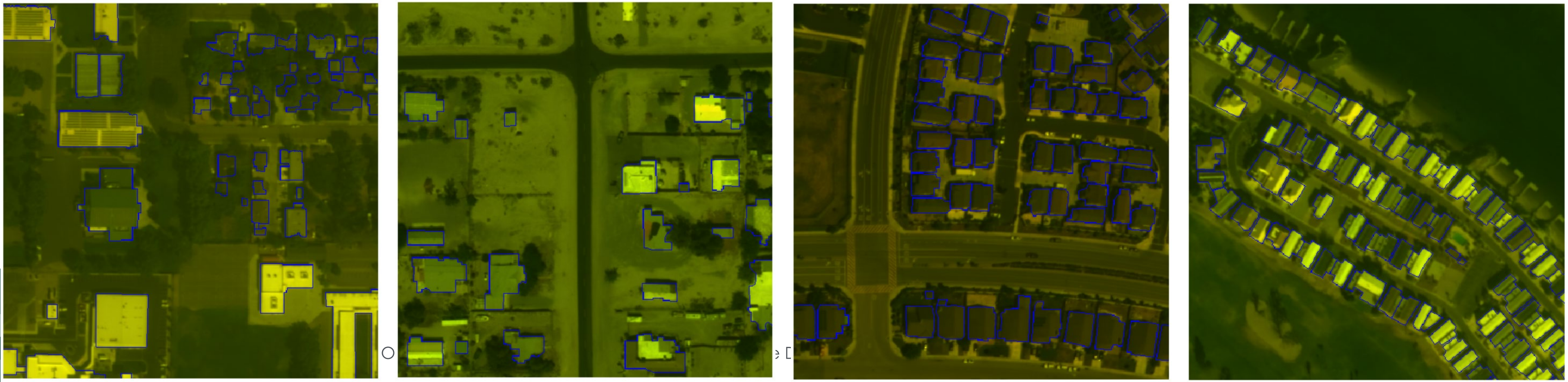


Development of a Robust & Generalizable U.S. Model

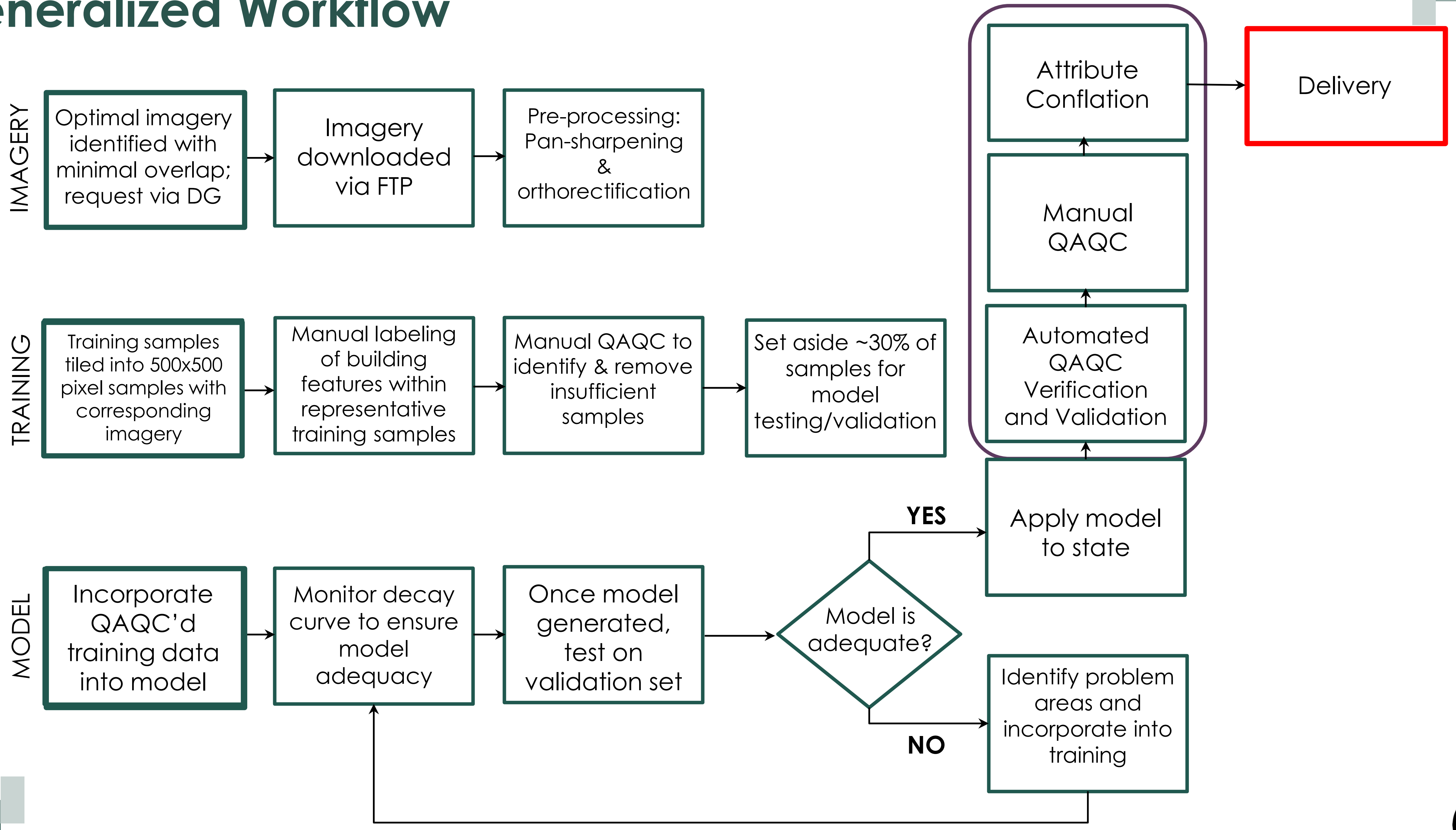
59,000+ manually created training samples:

- 25,500 positive samples
- 33,500 negative samples (no buildings present – aids in limiting false positive detections)





Generalized Workflow



Washington

Pennsylvania

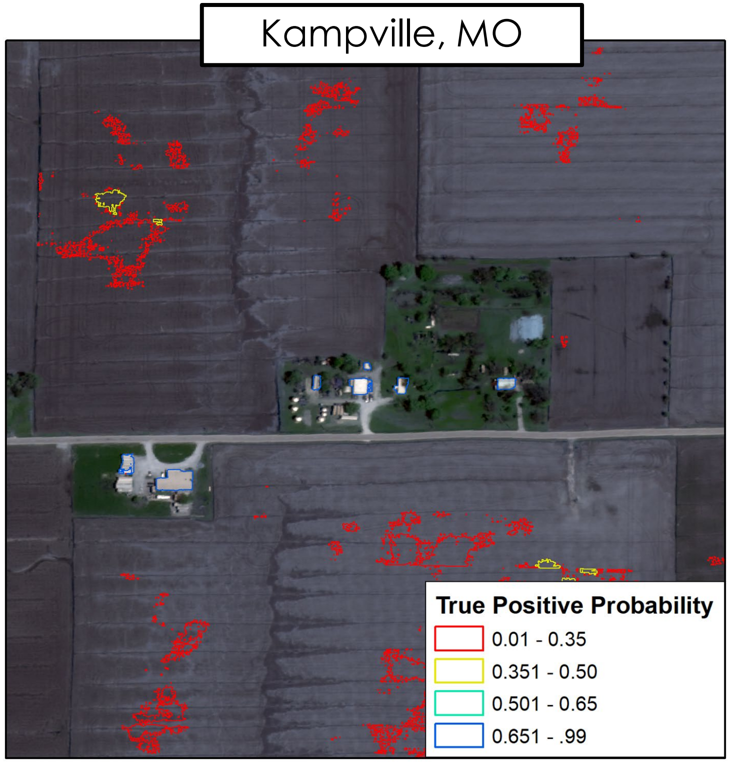
Florida

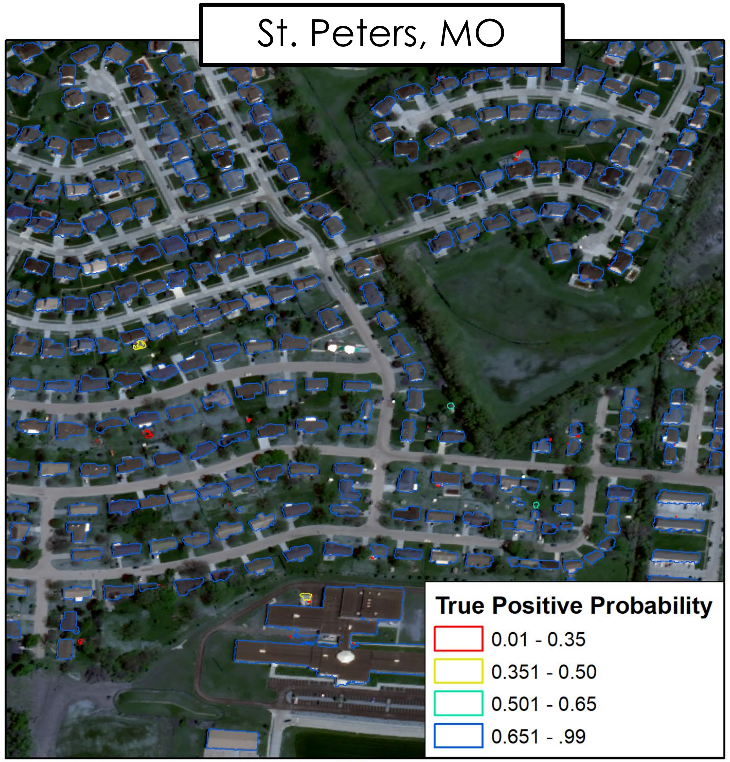
Colorado

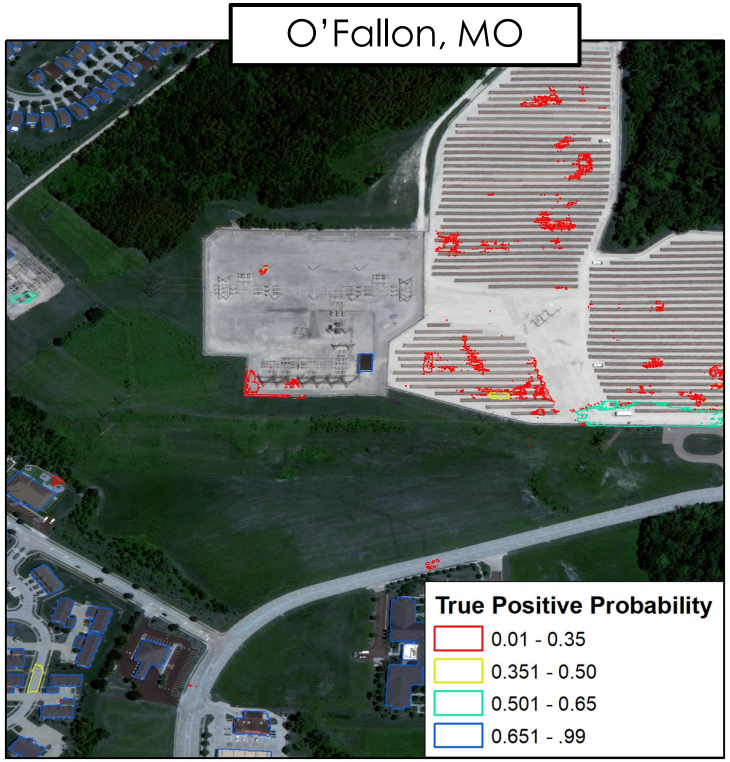
What is the Verification and Validation Model (VVM)

- A quality assurance and quality control model (QA/QC)
  - Designed to evaluate vectorized structure (building) detections
  - Identify and remove Type I Error (False Positive)
- Supervised Binary Classification Ensemble
  - Gradient Boosted Decision Tree

VVM Results





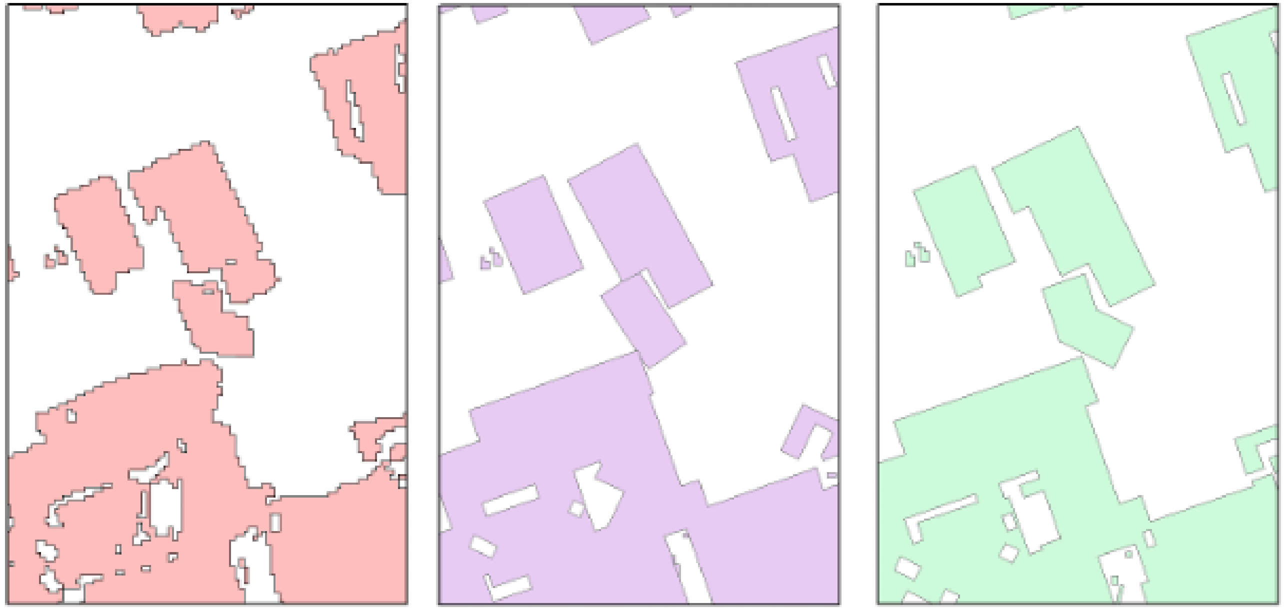


Sample Performance Results:

AOI	Correctly Labelled Features	Total Features	Accuracy
Kampville, MO	1,096	1,097	99.91%
St. Peters, MO	319	319	100.00%
O'Fallon, MO	718	720	99.72%

Regularization

- Benefits
  - Removes non-incident vertices
  - Improves render speeds in GIS software
  - Greatly reduces storage requirements
  - Converts detections into more familiar shapes
- Drawbacks
  - Computationally expensive
  - Can cause difficult to find issues
  - Another algorithm between imagery and final output



Occupancy Class – General Workflow

HIFLD
- Label Structures that intersect with HIFLD Layers

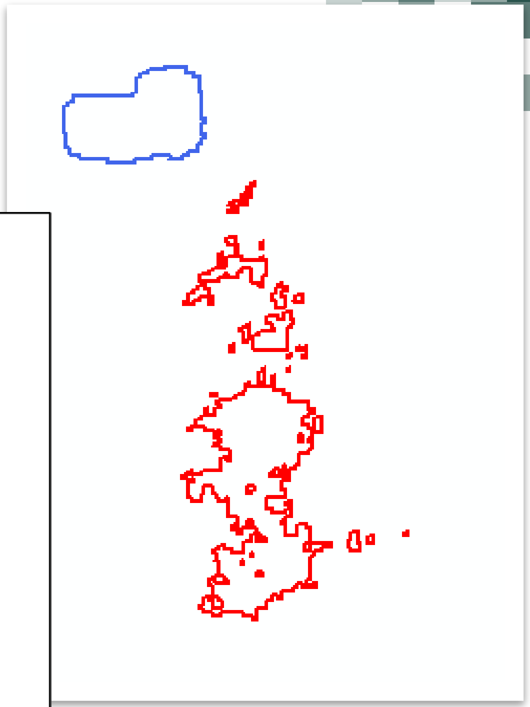
Census HU or Lightbox Parcels
- Remaining Structures then labeled with either Census or Parcels

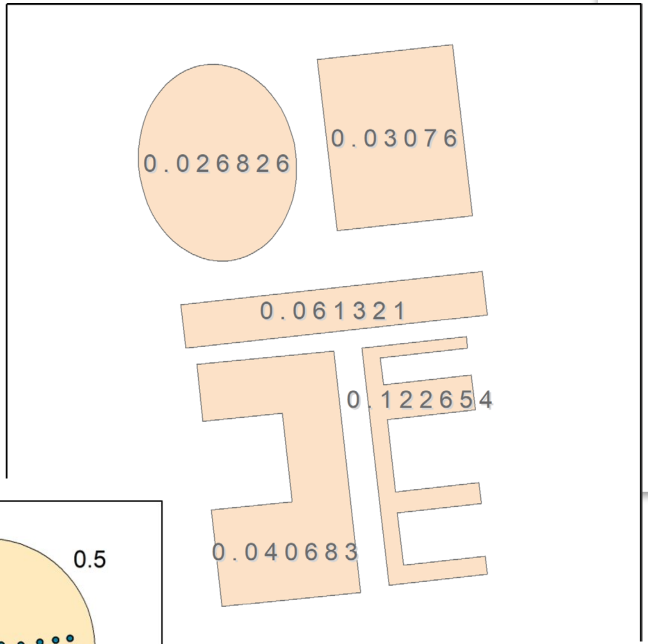
ResType
- Fill gaps of unlabeled structures with ResType

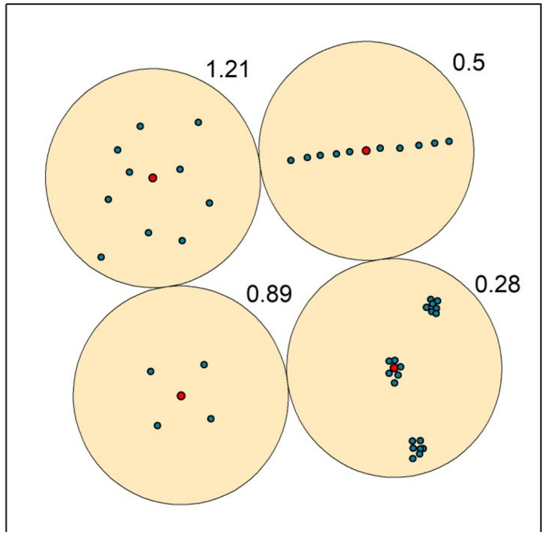
Occupancy Class Domain
- Residential
- Commercial
- Industrial
- Agriculture
- Assembly
- Non-Profit
- Government
- Education
- Utility and Miscellaneous
- Unclassified

GAUNTLET - A Tool for Calculating Building Morphologies

- Geometric
  - Measures of Geometry
- Engineered
  - Complex measures generated using geometric features
- Contextual
  - Spatial Point Patterns
  - Scale Relationships



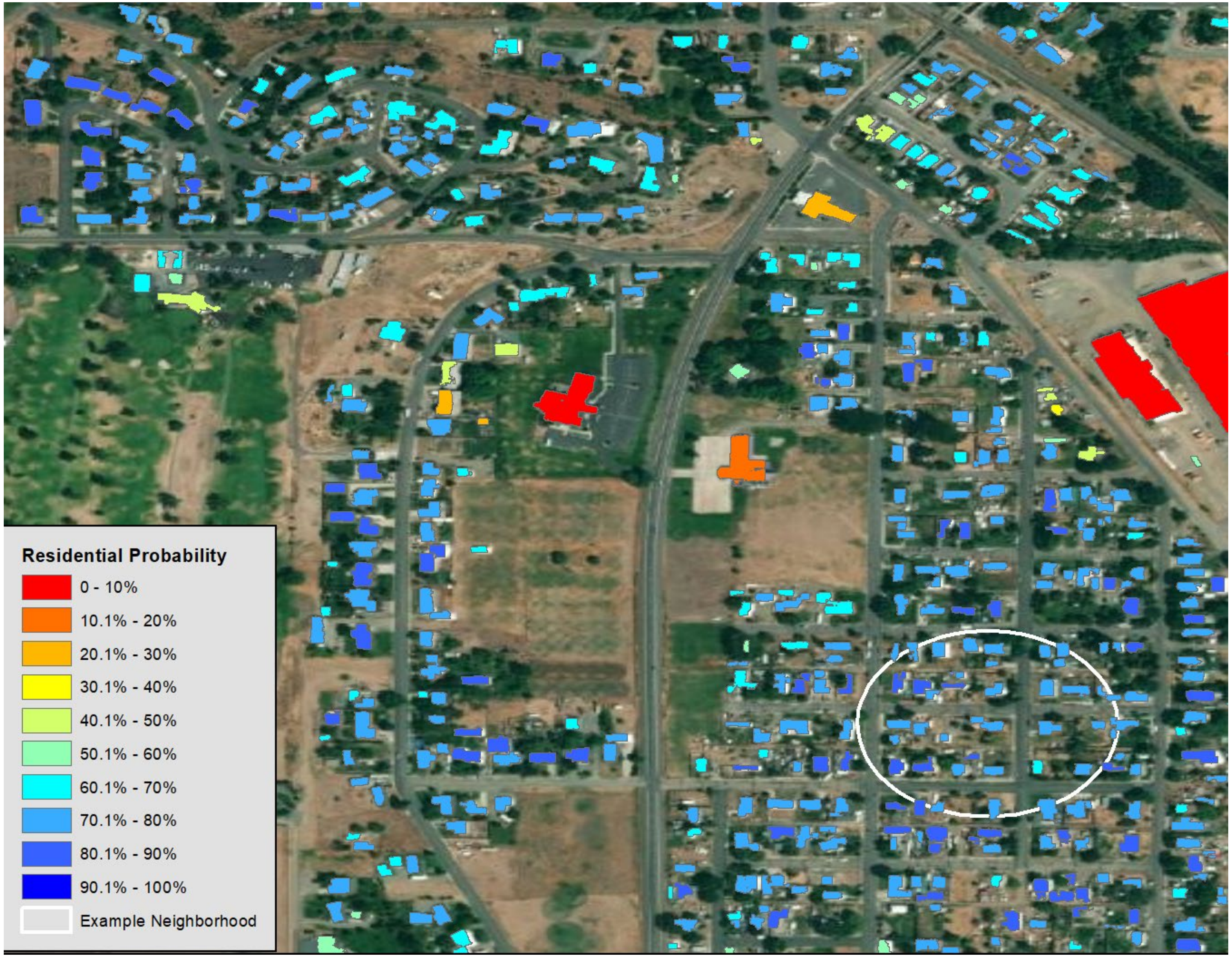




ResType Model

- Binary classification of building use
  - Residential
  - Non-Residential
- Analyzes building morphologies (Gauntlet features)
- Meant to fill a data gap
- Trained on labels from parcel data

Results



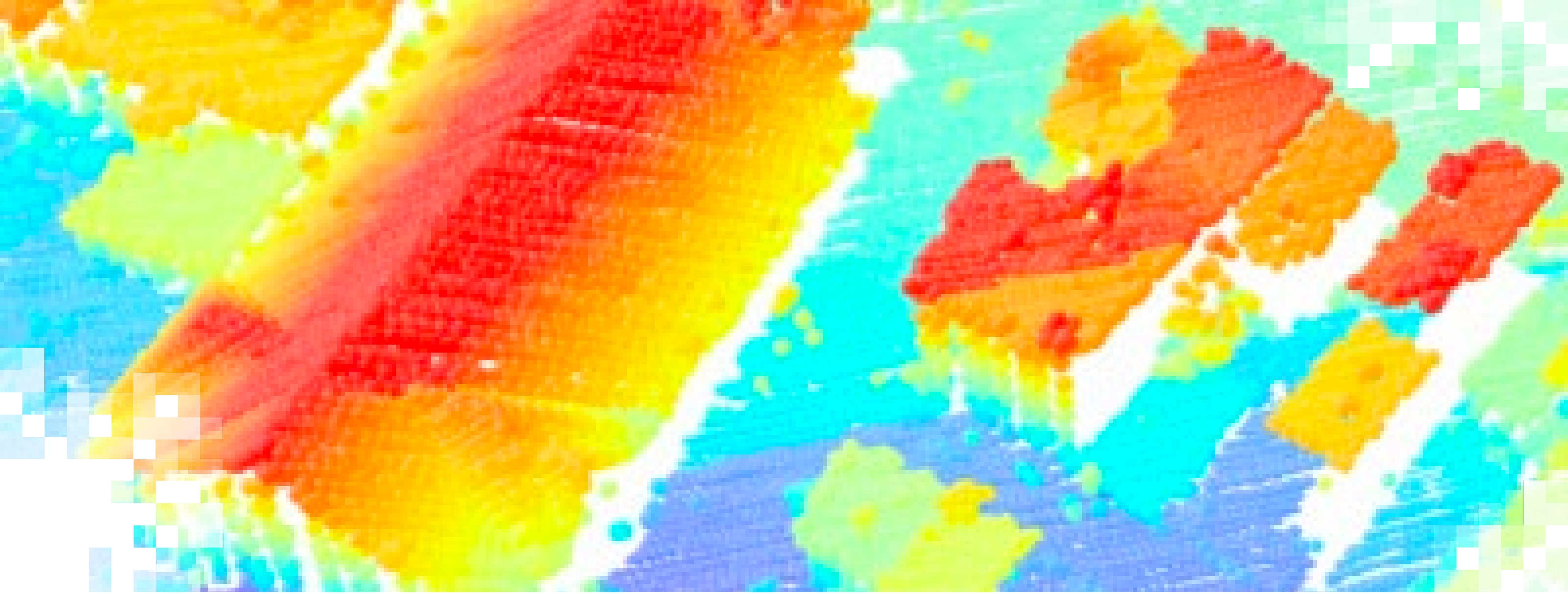
Residential Probability
0 - 10%
10.1% - 20%
20.1% - 30%
30.1% - 40%
40.1% - 50%
50.1% - 60%
60.1% - 70%
70.1% - 80%
80.1% - 90%
90.1% - 100%
Example Neighborhood

Bringing the Models Together

- Imagery processing
- CNN feature extraction
- VVM
- Regularization
- Attribute Workflow
  - 60+ datasets conflated
  - Gauntlet
  - ResType

Thank you

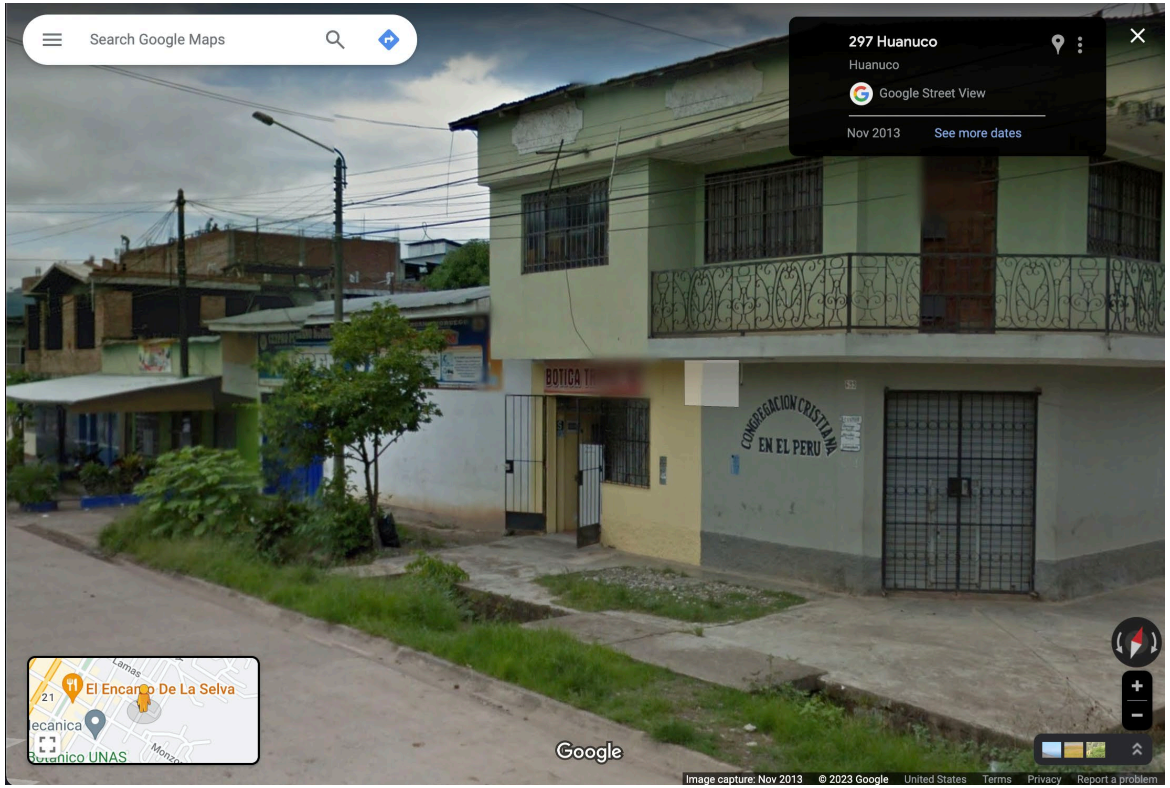
Taylor Hauser
hausertr@ornl.gov



Sampling from Google StreetView to Characterize Vulnerability

Sampling from Google StreetView to Characterize Vulnerability

- For areas without detailed infrastructure inventories, can we estimate vulnerability from samples?
- Using broad categories from remote sensing classification (urban core, industrial areas, residential areas), we can sample locations in StreetView to assess vulnerability.



Training Learning Objectives

By the end of this training, participants will be able to:

- Understand techniques for generating a random geographic sample
- Construct a survey of geographic components

Estimating Building Characteristics with Google StreetView

- Area-based estimates of building vulnerabilities are relevant for disaster planning and scoping exercises (decided where to collect more data).
- A random sample of urban areas can be used to estimate prevalence of vulnerabilities and to characterize different land use categories.

Approach

1. Survey design
2. Select sample sites
3. Setup Mechanical Turk
4. Data collection and analysis

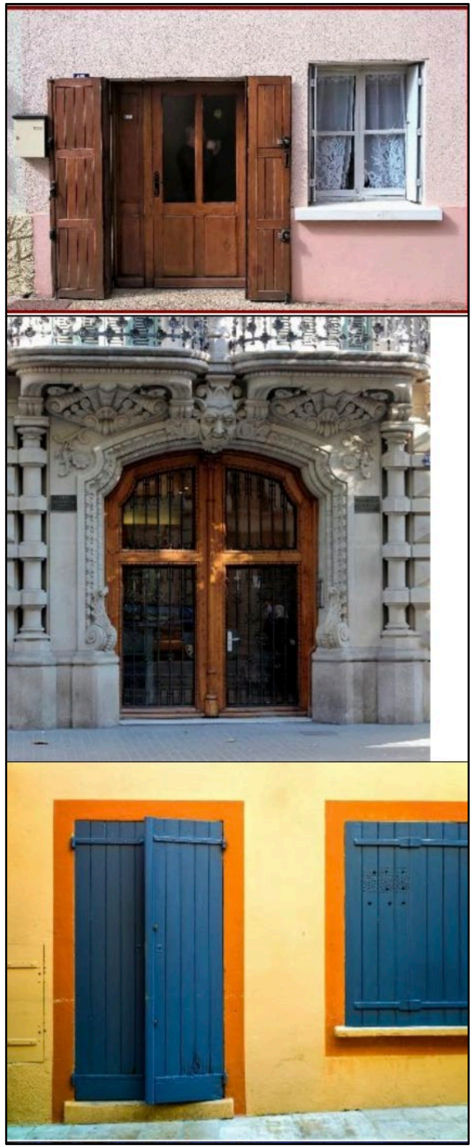
Survey Design

- Define variables
- Standardize through codebook
- Develop template task script of survey variables

Variable	Entry Options
Building Material	Masonry, cinder block, wood, brick, steel
Roof Type	Corrugated metal, tile, thatched or pal leaves, tar, gravel, concrete, mixed
Street Type	Asphalt, cobble or cement, dirt or gravel, potholed
Land Use	Residential, Commercial, Natural, Agricultural, mixed
Structure Type	Detached, semi-detached, attached, not applicable
Occupancy status	Occupied, vacant, cannot determine occupancy
Still Height (lowest entry point)	Ground level, low (1-6”), medium (7-12”), high (12-18”)
Building Condition	Very poor, poor, fair, good with minor defects, very good
Street Topography	Flat or low slope, medium slope, steep slope
Floors	Number of floors
Drains	Number of drains (street)

Selection of Sample Sites

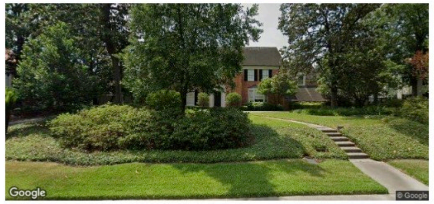
- Create random points within urban area.
- Locate buildings near points (drop points if no buildings present).
- Ensure that the buildings are included in Google StreetView.



Set Up Mechanical Turk

Mechanical Turk is a task-based system for repetitive work, such as reviewing images or other data.

- Export remote sensing optical images for each point and load in web accessible folder (e.g., Amazon Web Services "bucket").
- Install Mechanical Turk on local machine.
  - Alternatively, use Amazon paid service.
- Run module with template task script.



Building / Structure Information

Select a Sill Height

○ None, Ground Level ○ Low, 1-6" ○ Medium, 7-12" ○ High, 12-18" ○ Not Applicable

☐ Detached ☐ Semi-detached ☐ Attached ☐ Not Applicable

[1] Number of Floors

Building Condition, Status and Material

○ Very Poor ○ Poor ○ Fair ○ Good with Minor Defects ○ Very Good

☐ Under Construction ☐ Masonry or Cinder Block ☐ Wood Construction ☐ Brick ☐ Steel

○ Occupied ○ Vacant ○ Cannot determine occupancy

Roof Type

☐ Corrugated Metal ☐ Tile ☐ Thatched or Palm Leaves ☐ Tar ☐ Gravel ☐ Other: [Additional Notes]

Land Use

☐ Residential ☐ Commercial ☐ Agricultural ☐ Industrial ☐ Natural ☐ Other: [ ]

Street Information

○ Flat or Low Slope ○ Medium Slope ○ Steep Slope

[0] Number of Drains Visible

☐ Paved Street (Asphalt) ☐ Cobble or Cement Blocks ☐ Dirt or Gravel ☐ Potholes ☐ Other: [Additional Notes]

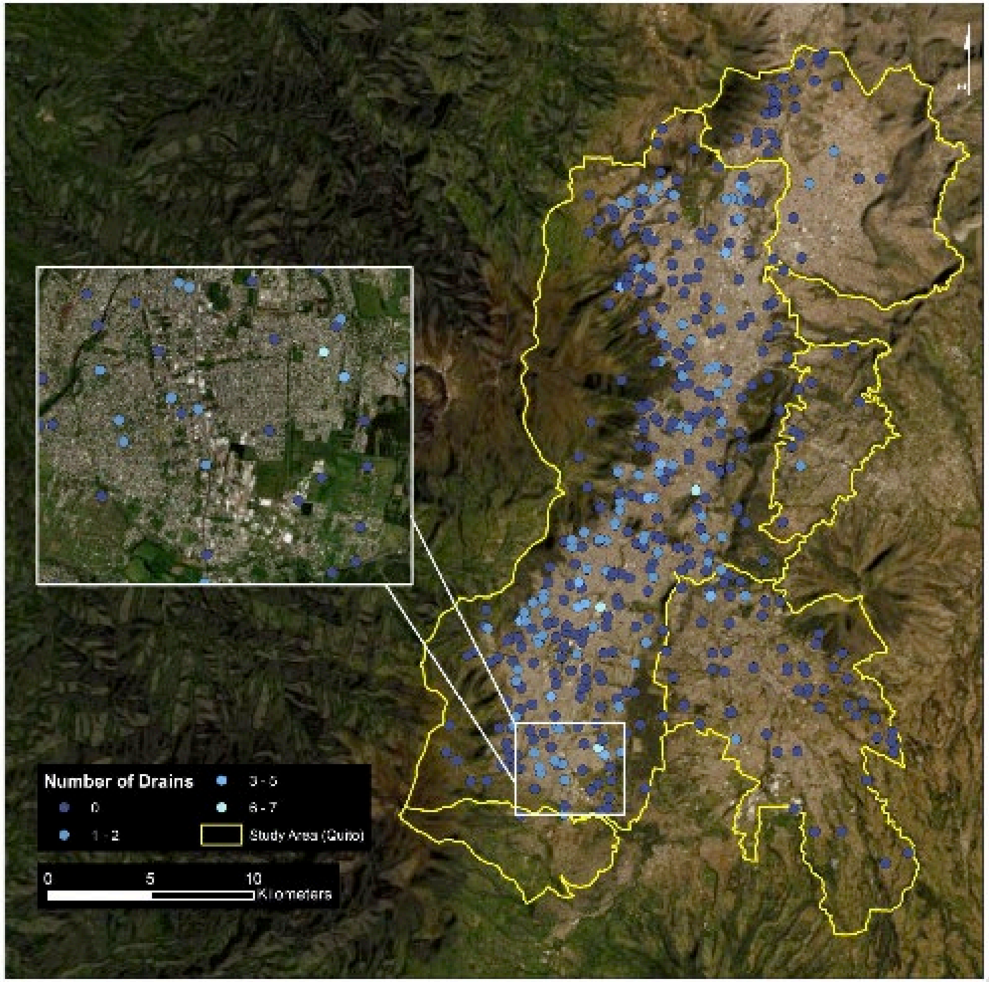
[Add any notes here]

☐ Google StreetView Not Available

[Submit]

Data Collection and Analysis

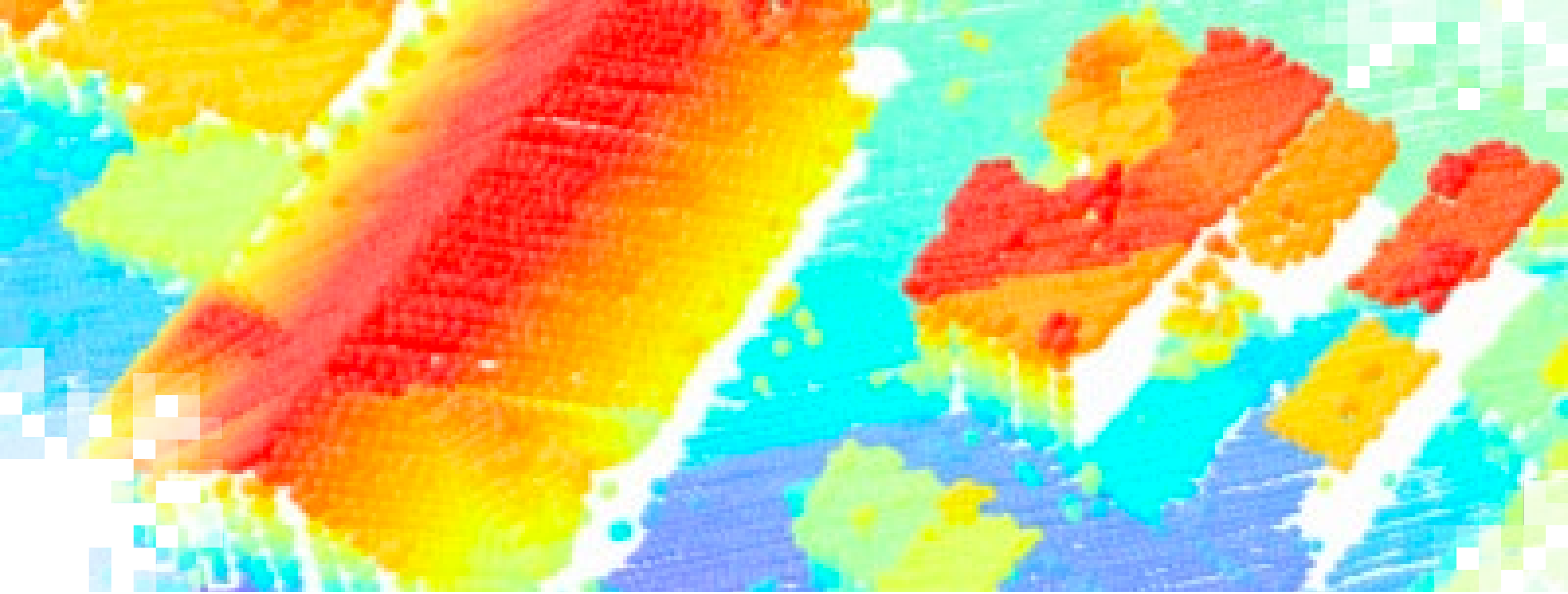
- Use Mechanical Turk to collect data for each location.
- Evaluate data collected for each variable.
  - Review consistency across analysts.
- Visualize results in context.



Results and Conclusions

- Flood vulnerability variables for hundreds of buildings can be collected in just a few hours.
- Uncertainty of output should be measured.
  - Easy for different analysts to interpret instructions differently; examples are useful
  - Not always consistent between analysts

Variable	Kappa (κ)
Building Material	0.94
Roof Type	0.83
Street Type	0.87
Land Use	0.88
Structure Type	0.64
Occupancy Status	0.80
Sill Height	0.66
Building Condition	0.54
Street Topography	0.79



Part 2:
Summary

Summary

- Developing a building-level exposure data set for HAZUS Flood Study in New York
- Using Earth Observations to develop a building structures dataset
- Case study: Sampling from streetview to characterize vulnerability

Looking Ahead

- Part 3: Development of Site-Specific Exposure Data with EO
  - Exposure data best practices
  - Developing and understanding metadata
  - Equity and bias considerations
  - Case study: Assessing climate change impacts with building exposure data in Antigua and Barbuda

Contact Information

Trainers:

- Greg Yetman
  - gyetman@ciesin.columbia.edu
- Taylor Hauser
  - hausertr@ornl.gov

- ARSET Website
- Follow us on Twitter!
  - @NASAARSET
- ARSET YouTube

Visit our Sister Programs:

- DEVELOP
- SERVIR

Resources

- New York State building footprint data and Mapper
- Amazon Mechanical Turk
- Local Mechanical Turk (open source)

Thank You!