Integrated Spatio-Temporal Depictions of the Built Environment for Improved Understanding of Settlement Changes

Yoonjung Ahn[1], Johannes Uhl[1&2], Stefan Leyk[1]
Yoonjung.Ahn@colorado.edu

[1] Institute of Behavioral Science, University of Colorado, Boulder, CO, USA
[2] European Commission Joint Research Centre, Ispra, Va, Italy

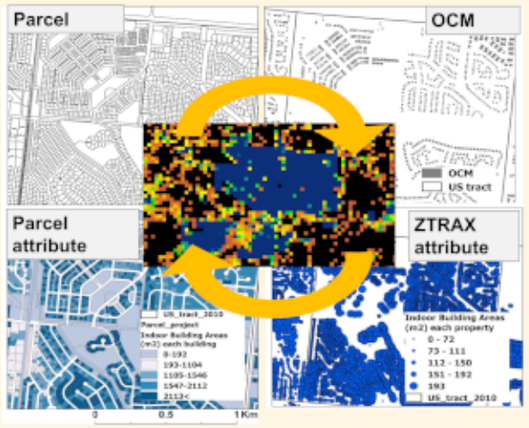
1. Background & Objectives

The knowledge iof the evolution of human settlements is crucial for understanding environmental change and human-environment interactions. In recent years, various spatial data products have been developed that describe the built environment at national, regional, and global levels, such as Google's building footprints, OpenStreetMap, and Open City Model (OCM) data. While these datasets capture the human footprint in unique ways, developing comprehensive historical settlement data has been challenging due to data missingness and infrequent data updates.

We address this issue by integrating various sources of large-scale geospatial data. These efforts were made to improve coverage of our recent work on the Historical Settlement Data Compilation for the US (HISDAC-US)[1&2]

2. Data & Methods

We use **Zillow's ZTRAX**[3] database, containing ~300 million data records on property & building characteristics and real estate transactions in the US, **OpenCityModel**[4] and Microsoft's building footprint data, and **Parcel data from Parcel Atlas**[5], including ~146 million current parcels in the US. Integrating the described spatial input data is challenging because the data is heterogeneous in its basic properties and quality parameters. This processing step aims to create a higher-quality combined data product using the different characteristics of the input data sources. In this case, data integration is based on various spatial operations, such as spatial join (intersection, nearest join) and address matching, to create a dataset of maximum completeness featuring building attributes such as lot size, building indoor area, volume, and year-built. After integrating ZTRAX, Parcel, and OCM data, we converted them to point features. The point features were input to the rasterization step. We created a set of multi-temporal gridded layers using the built-year attribute for temporal definition and building locations and attributes within grid cell extents to calculate and assign the cell values.

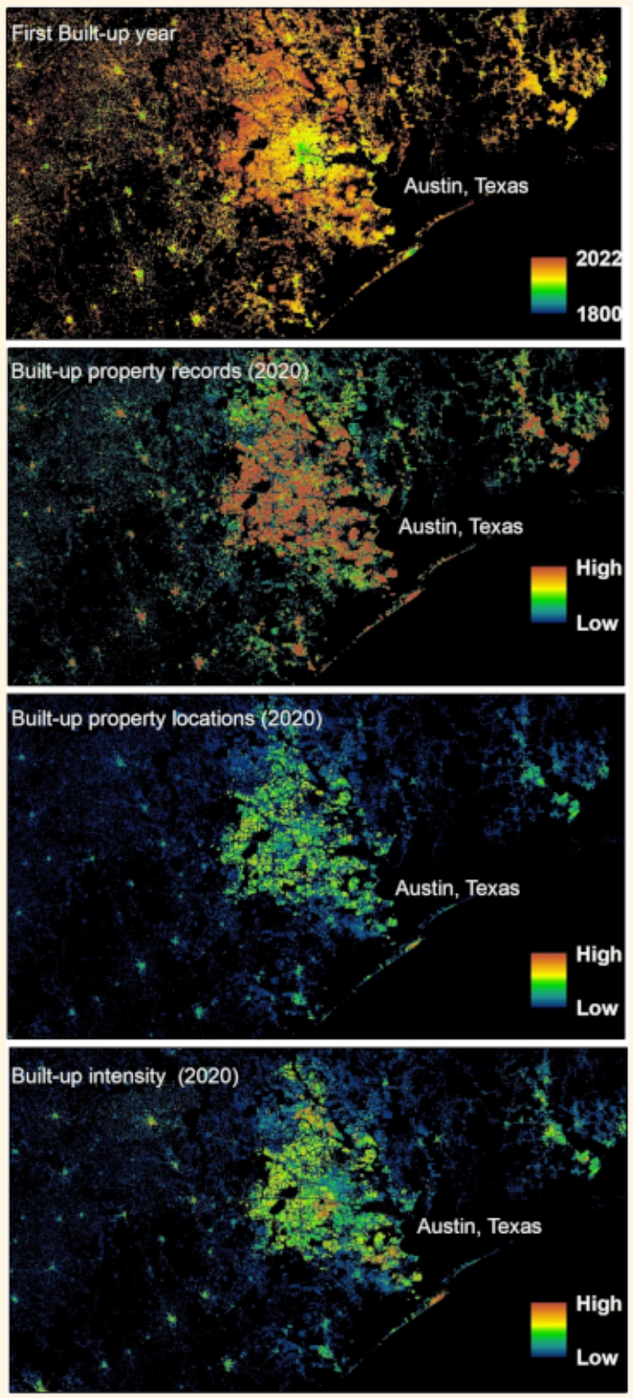


Data	HISDAC version1	HISDAC version2	Data format
Zillow's ZTRAX	O	O	points
Parcel		O	polygons
Building footprint		O	polygons

3. Results and Discussion

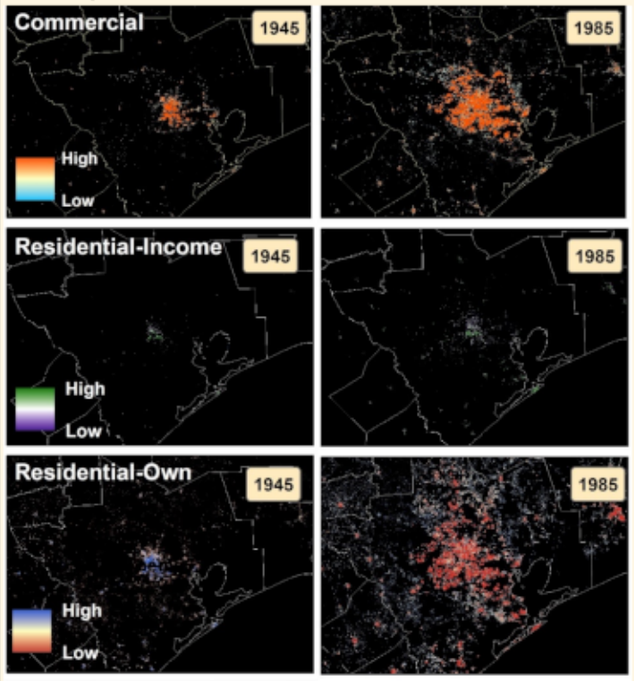
3.1 Built up areas

We created 250m resolution gridded surfaces of First Built-up year (FBUY), Built-up property locations (BUPL) Built-up property records (BUPR), and Built-up intensity (BUI).



3.2 Land use

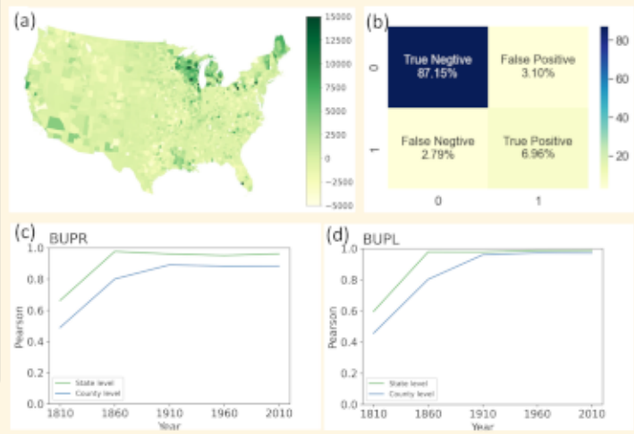
We also created 250m resolution gridded surfaces of 8 land use categories, including commercial, residential income, residential own, agricultural, recreational, industrial, governmental, and vacant from 1940 to 2020. The figure below presents the surfaces for different land use categories (commercial, residential income, and own) in 1945 and 1985 in Austin Texas.



3.3 Comparison

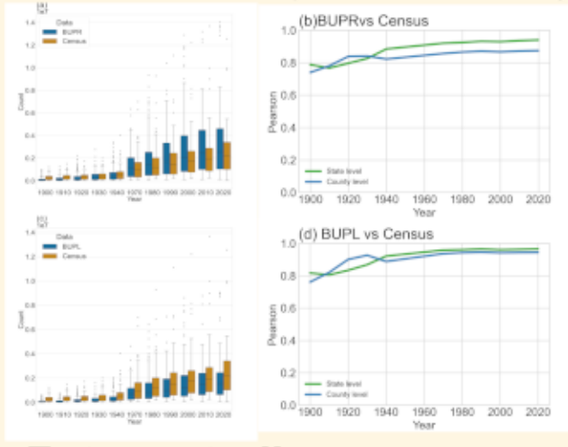
3.3.1. HISDAC v1 VS **HISDAC v2**

The results in below figure show that HISDCAv2 has higher completeness of year-built information particularly in the Midwest and is strongly correlated to HSDAC v1 after the1860s. Fig. (a) illustrates year-built coverage (v2-v1); Figs. (c) & (d) show the Pearson correlation correlation between the two versions of BUPL & BUPR at the state and county level over time.



3.3.2. Census vs **HISDAC v2**

The figures below illustrate the similarity between US-wide housing trends (1890s - 2020s) when using U.S. Census and HISDAV_v2 BUPR data at the county and state level. Pearson correlation ranged from 0.75 to 0.9. These results are reasonable considering HISDAC_v2 BUPR data include all land use classes while census data only contains residential housing.



4. Future studies

We will continue to conduct comparison analysis for land use data. These spatial data collections will be further expanded by combining different input data sources to derive additional variables such as building volume used for natural hazards exposure assessments.

5.Reference

1. Leyk S. and Uhl J.H. 2018. HISDAC-US, Historical Settlement Data Compilation for the Conterminous United States Over 200 Years. Nature Scientific Data 5:180175, https://doi.org/10.1038/sdata.2018.175
2. Uhl JH., Leyk S., McShane CM., Braswell AE., Connor DS., and Balk D.2021. Fine-grained, spatio-temporal datasets measuring 200 years of land development in the United States, Earth System Science Data 13:119-153, https://doi.org/10.5194/essd-2020-217
3. Zillow. (2022). *Zillow's Assessor and Real Estate Database*. https://www.zillow.com/research/ztrax/
4. BuildZero. (2019). *Open City Model*. https://github.com/opencitymodel/opencitymodel#license
5. Parcel Atlas. (2022). *Parcel*. https://www.boundarysolutions.com/BSI/ParcelAtlas/page1.html

6.Acknowledgments

Funding for this work was provided through the Humans, Disasters, and the Built Environment and the Human Networks and Data Science – Infrastructure programs of the US National Science Foundation (Award Numbers 1924670 and 2121976, respectively) to the University of Colorado Boulder. This research benefited from support provided to the University of Colorado Population Center (CUPC, Project 2P2CHD066613-06) from the Eunice Kennedy Shriver Institute of Child Health, Human and Human Development. The content is solely the authors' responsibility and does not necessarily represent the official views of NIH or CUPC.

University of Colorado Boulder
IBS INSTITUTE OF BEHAVIORAL SCIENCE

Google Scholar
YoonJung Ahn - Google Scholar
Yoonjung Ahn, PhD (she/her) (@lostgeographer...