Creating a Generalizable Model for Urban Heat Islands in the United States

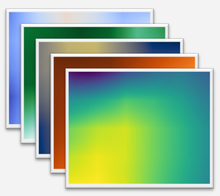
Yazlin Moujalled[3], Oumaima Berrada[2], Rittik Barua[2], Ella Tallett[1], Zach Calhoun[1], David Carlson[1], Mike Bergin[1]
Duke University[1], Duke Kunshan University[2], University of Maryland Baltimore County[3]

Abstract

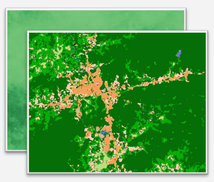
Urban heat islands (UHIs) - areas within cities that are significantly hotter than surrounding rural environments due to urban morphology - are widespread across the United States. Elevated urban temperatures increase electricity demand and heat-related illnesses, placing substantial stress on energy and public health infrastructure. Understanding **how climatic conditions and land cover contribute to urban heat** is critical for developing effective adaptation and mitigation strategies. While many UHI models have been developed on a city-by-city basis, building a model capable of predicting heat across all U.S. cities represents a crucial step toward anticipating future heat impacts, especially in data-sparse regions. In this study, we primarily **investigated two modeling approaches** to predict urban heat using land cover and climatic data: linear regression with interaction terms and mixed effects model. We found potential in creating a generalizable model to predict urban heating, especially if more nuance is added into the model.

Methodology

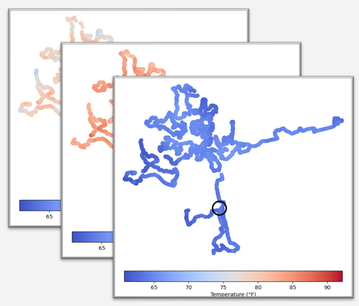
Data Acquisition and Alignment → **Data Pre-processing** → **Model Training and Evaluation**



ERA5-Land Data
~9km resolution
Wind Speed, Surface Pressure, Evaporation, Solar Radiation, Dew Point



LCZ (land-use) and NDVI Data
100m | 30m resolution
using Google Earth Engine



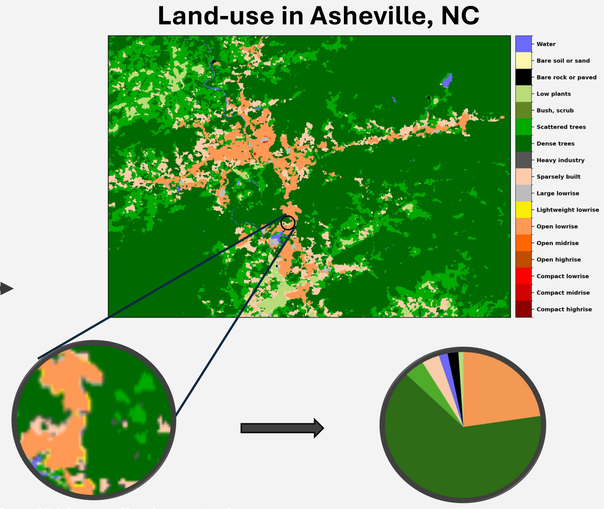
NOAA Temperature Data
From 3 different time slots: morning, afternoon, & evening

Mapping the coordinates of NOAA data collection sites

Feature Engineering

- **ERA5-Land variables** — averaged per city for climate background
- **NOAA readings** — urban temp anomaly = reading$_{i,j}$ − mean(5th percentile$_{,i}$)
  where *i* indexes city and time slot, and *j* indexes individual readings
- **Averaged LCZ** — performed log-ratio transformation followed by Principal Component Analysis (PCA)

Land-use in Asheville, NC



Buffers (500m radius) created around NOAA data collection coordinates

Land-use composition computed for each buffer

Multi-city dataset per time slot
Data from 42 U.S. cites

Features
- Pre-processed land-use variables
- Averaged ERA5-land variables

Target
Temp. anomaly due to urban effects

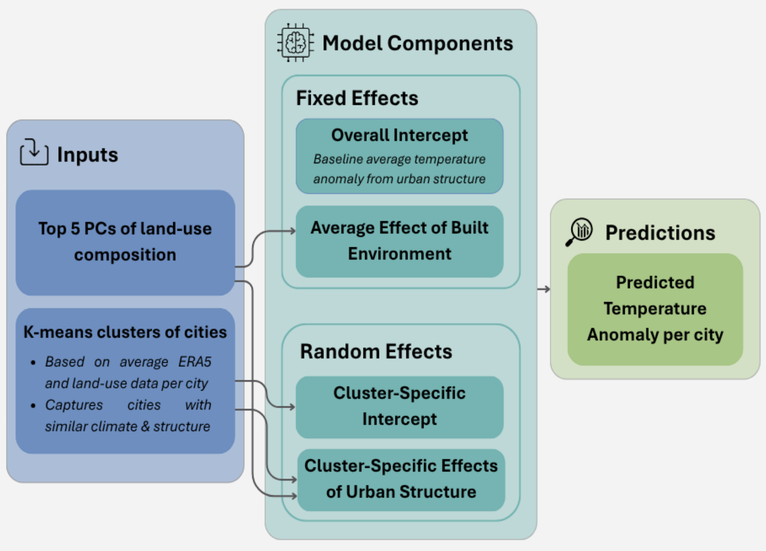
Data splitting

Randomly selected a subset of 32 cities → Training dataset → **Model training**
- Mixed Effects Model
- Linear Regression w/ Interactions

Held-out subset of 10 cities → Testing dataset → ***Predictions***
*Performance measured by **Pearson's r***

Modeling

Mixed Effects Model



Linear Regression with Interactions

Predicting temperature anomalies using raw land-use variables at 100m resolution. Interactions between ERA5 and land-use variables. Ridge regularization was applied to prevent overfitting.

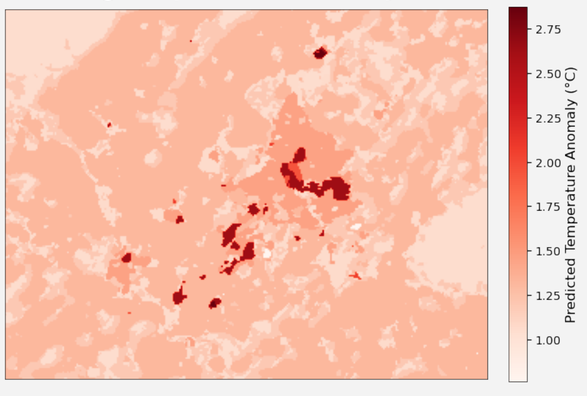
Table: Average Pearson's R on training and testing data

Model	Training data			Testing data		
	AM	AF	PM	AM	AF	PM
Mixed Effects Model	0.41	0.28	0.39	0.48	0.10	0.31
Linear Regression with Interactions	0.44	0.55	0.58	0.65	0.04	0.06

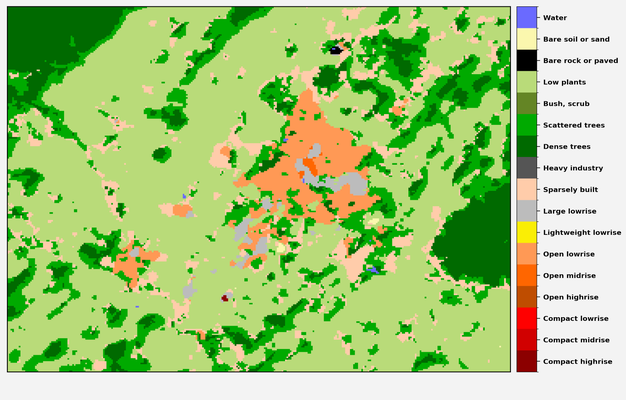
Results & Discussion

The results led to the following insights:
- The models reveal a qualitative positive relationship between urban structure and temperature anomalies
- The performance of the mixed effects model compared to the linear regression model shows the mixed effects model may be more temporally generalizable
- The Pearson's R for the AM linear model raises a point about land-use buffers

Harrisonburg, VA
Morning Temperature Anomaly Prediction



Land-use in Harrisonburg, VA



Figures for Harrisonburg, VA: land-use and temperature anomaly images for held-out city using *Linear Regression with Features*

Next Steps

Our next steps include:
- Implementing domain adaptation to improve generalization across cities
- Identifying more data sources, especially those with detailed urban structure information
- Incorporating a buffer to account for landcover influences on urban heat based on distance from reading
- Exploring more complex modeling techniques