Using Spatial Causal Inference to Model the Urban Heat Island Effect

Edrian Paul Liao[1,2]

[1] Electrical and Computer Engineering, Duke University, [2] Computer Science, Duke University

1 Urban heat island effect and its consequences

The urban heat island effect, one of the unfortunate consequences of the human-induced climate change, is characterized by rising temperatures in urban areas due to the **general lack of vegetation** and **albedo**–the fraction of light that a surface reflects. The associated thermal stress will directly impact **cardiovascular and respiratory morbidity and mortality**, **disproportionately affecting the urban poor**. It also **increases cooling energy demand** that in turn, increases carbon emissions.

2 Cool roof and urban greening strategies

The most common mitigation strategies have been utilizing (1) cool roofs, which is the application of a highly reflective coating to **increase albedo**, and (2) urban greening, which **increases the amount of vegetation**. Because such interventions highly depend on **location and the extent of application**, most studies on mitigating the urban heat island effect are **localized**.

3 The spatial causal inference model

We then propose an **interpretable yet accurate** urban heat island model that can be generalized across cities with **different geographic features**.

$$Y_{ij} = \gamma \mathbf{X_{ij}} + \beta_{15} V_{ij} + \beta_{16} \overline{V_{ij}} + \beta_{17} A_{ij} + \beta_{18} \overline{A_{ij}} + U_{ij} + \epsilon_{ij}, \gamma = (\beta_0, \beta_1, ..., \beta_{14}) \quad (1)$$

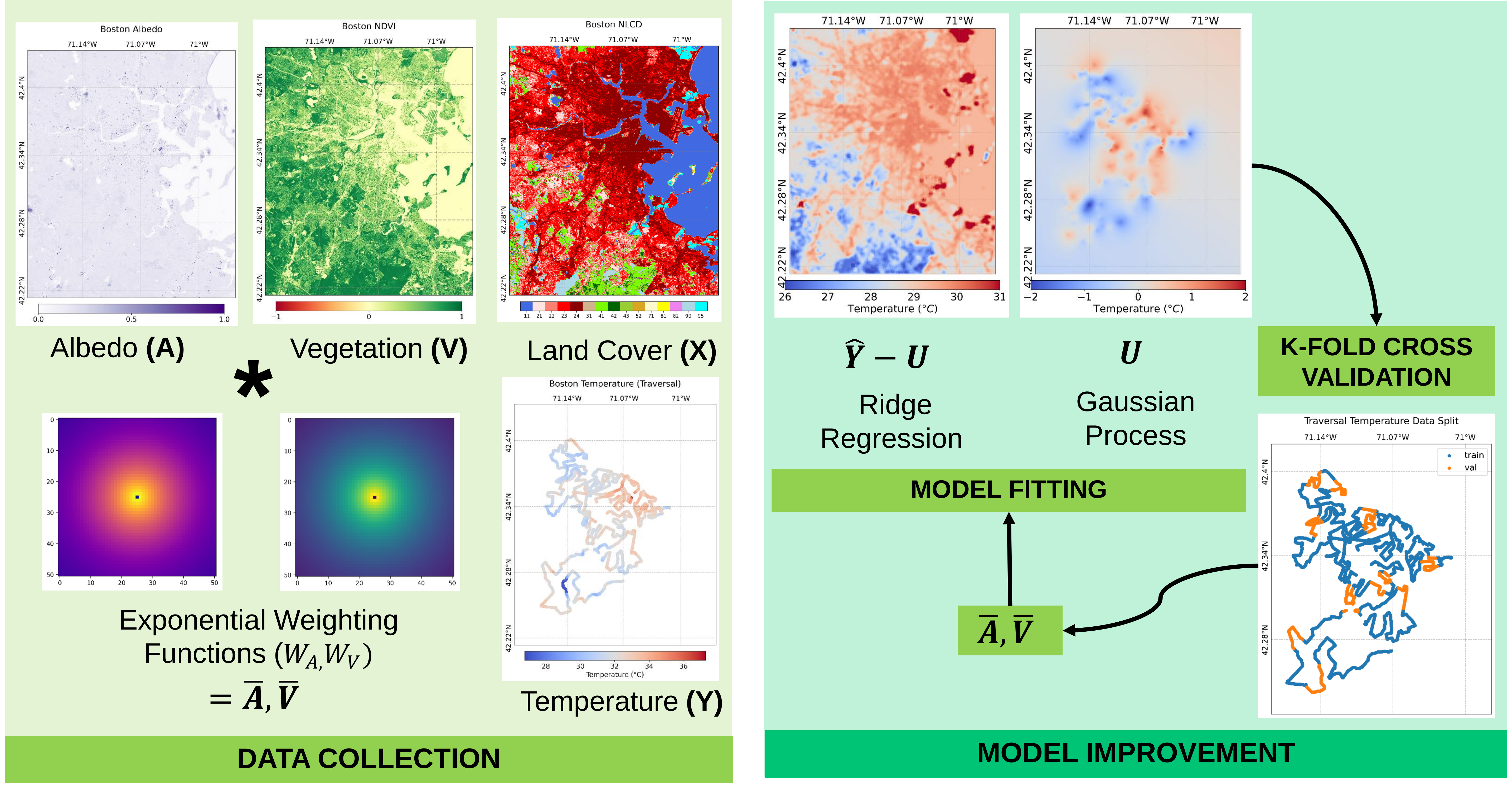
$$\bar{V_{ij}} = \sum_{k \neq 0} \sum_{l \neq 0} w_V(\sqrt{l^2 + k^2}) V_{i+k,j+l} \quad (2)$$

$$\bar{A_{ij}} = \sum_{k \neq 0} \sum_{l \neq 0} w_A(\sqrt{l^2 + k^2}) A_{i+k,j+l} \quad (3)$$

Eq. 1 estimates the linear causal effect of treatment described by the changes in **Normalized Difference Vegetation Index (NDVI)** ($\boldsymbol{V_{ij}}$) and **Albedo** ($\boldsymbol{A_{ij}}$) data from Sentinel-2 satellite images on an outcome defined by the **air temperature data** ($\boldsymbol{Y_{ij}}$) from NOAA. The **vector** $\boldsymbol{X_{ij}}$ represents the **observed confounders**, i.e., the 15 land classes.

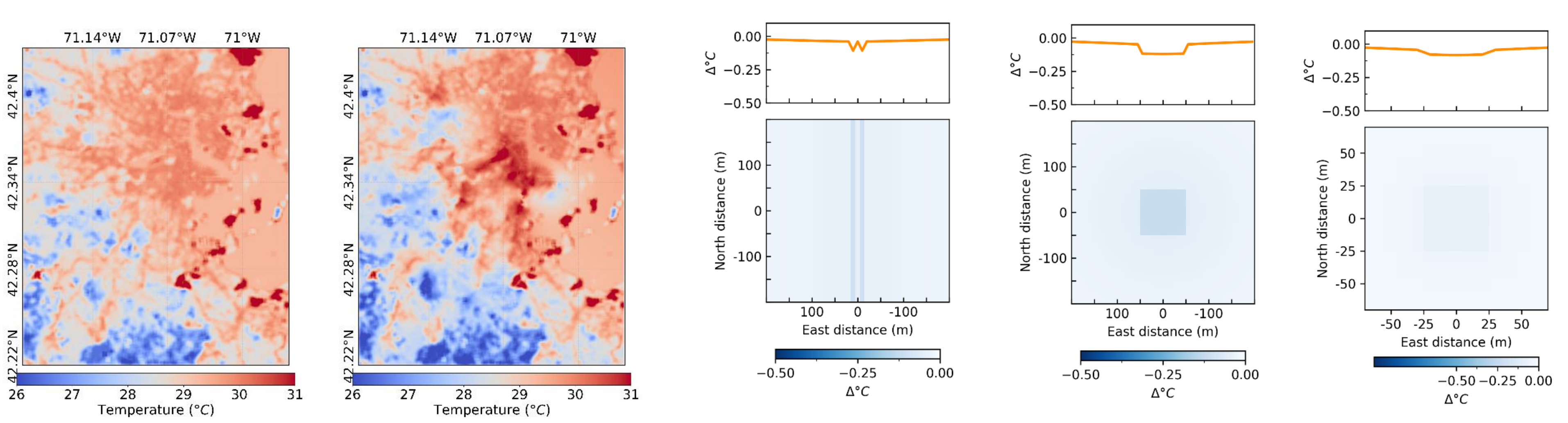
Eq. 2-3 represent the **spatial terms for NDVI and Albedo** that are subjected to two different exponential weighting functions to consider the varying gravity of neighboring effects, depending on the Euclidean distance of two points (i, j) and $(i + k, j + l)$.

4 Data collection, preprocessing, and training the model



5 Using the learned model to predict intervention effects

The optimal hyperparameters for the exponential weighting function were found to be length scales of **26 (260 m) for NDVI data** and **8 (80 m) for albedo data**. The **Ridge Regression R2 score** was **0.45** and the **combined Ridge-Gaussian process R2 score** on all of the data was **0.91**. With this model, we can predict the overall effect of current mitigation strategies on an urban area's air temperature.



Ridge Regression | Ridge Regression + Gaussian Process | Street with trees | Park in an urban block | Cool roof

6 Conclusion

This model could **provide recommendations to policymakers** on the optimal location of such interventions. This will be increasingly important as we try to adapt to a warming climate and mitigate said health risks and energy demands. Future research include (1) exploring the effects of blue infrastructure by taking into account a new water-related input variable, the **Net Difference Water Index (NDWI)** and (2) applying the same method to **more cities with different geographic features**.

Check out the codebase used to model and generate these visualizations!