PUSHING CARTOGRAPHIC BOUNDARIES: USING MACHINE LEARNING APPROACHES TO MODEL KEY PREDICTORS OF SOIL RESPIRATION IN MIDDLE TENNESSEE

Caroline Y. Apraku [1]*, Reginald S. Archer, PhD, GISP [1]

[1] Department of Environmental Science, Tennessee State University, Nashville, USA
* capraku@my.tnstate.edu

Keywords: cartography, Machine Learning, Middle Tennessee, pavement, soil respiration**,** urban heat

Urbanization, Urban Heat and its Impacts

Soil respiration is the main pathway that carbon moves from soil into the atmosphere and it tends to increase when temperatures rise - especially through heterotrophic respiration driven by microbial activity. Understanding how soil respiration interacts with the Urban Heat Island (UHI) effect is key to fully grasping this process. Instead of relying on traditional, static soil maps, machine learning models like XGBoost and SHAP enable Dynamic Carbon Mapping.

UHIs, characterized by significantly elevated air temperatures (Oke, 1982; Grimmond, 2007), presents a major challenge for urban management. UHIs have substantial implications for urban life, including agriculture. Elevated temperatures influence soil moisture dynamics, microbial activity, and plant physiological processes, which can reduce crop productivity and alter nutrient cycling in urban farms (Beniston et al., 2018). Increased soil temperatures accelerate organic matter decomposition, potentially reducing long-term soil carbon stocks and altering greenhouse gas emissions (Crowther et al., 2016). From a public health perspective, UHIs exacerbate heat stress, increase the risk of cardiovascular and respiratory diseases, and contribute to elevated mortality during extreme heat events (Harlan et al., 2013; Gasparrini et al., 2015). Nighttime heat retention limits recovery from daytime heat stress, disproportionately affecting vulnerable populations and raising environmental justice concerns (Hsu et al., 2021). Elevated energy demand for cooling further increases emissions and air pollution, creating feedback loops that worsen health risks.

Urban heat islands also influence the global carbon cycle through soil biogeochemical processes. Soil respiration, the primary pathway for soil carbon fluxes to the atmosphere, is enhanced under elevated temperatures, particularly heterotrophic respiration driven by microbial decomposition (Kaye et al., 2005; Zhou et al., 2021). Altered soil moisture regimes may influence methane production and oxidation, further contributing to atmospheric greenhouse gas concentrations (Treat et al., 2015).

Urbanization, which primarily drives the UHI therefore fundamentally alters the Earth's "living skin." In the rapidly expanding corridor of Middle Tennessee, soil is mostly treated as an inert substrate for development rather than a dynamic biological system.

However, urban soils act as critical carbon sinks and sources. The interaction between UHIs and Urban Dry Island (UDI) zones (lower soil moisture than their surroundings), creates microclimate stressors that are difficult to map traditionally. Soil respiration is also influenced by total organic carbon (TOC), total nitrogen (TN), and microbial activity. The results of Analysis of Covariance (ANCOVA) confirmed that other than moisture (p-value <0.001), TOC, TN, Land use and soil temperature were not strong predictors of respiration (p-value >0.05)

This study builds on initial findings of TOC and TN sensitivity by examining the interactions between each factor as potential predictors of carbon efflux.

For this study, we move from observation to prediction, using machine learning to map the "breath" of the city - soil respiration. The goal is to predict carbon efflux using only readily available environmental and geospatial predictors.

Figure 1 shows soil respiration is governed by two primary axes (Dim 1 and 2) where dimension 1 shows resource availability (moisture and nutrients) and dimension 2 shows clustering for microclimate conditions and thermal stressors. **Figure 2** shows that the relationship between soil moisture and respiration has a positive correlation where increase in one causes an increase in the other variable. Unlike this relationship, **Figure 3** shows a static relationship, where pavement does not directly have a strong influence on respiration. This lack of a direct linear relationship is further supported by the ANCOVA p-value (>0.05), confirming that pavement is not a statistically significant predictor when compared to soil moisture.

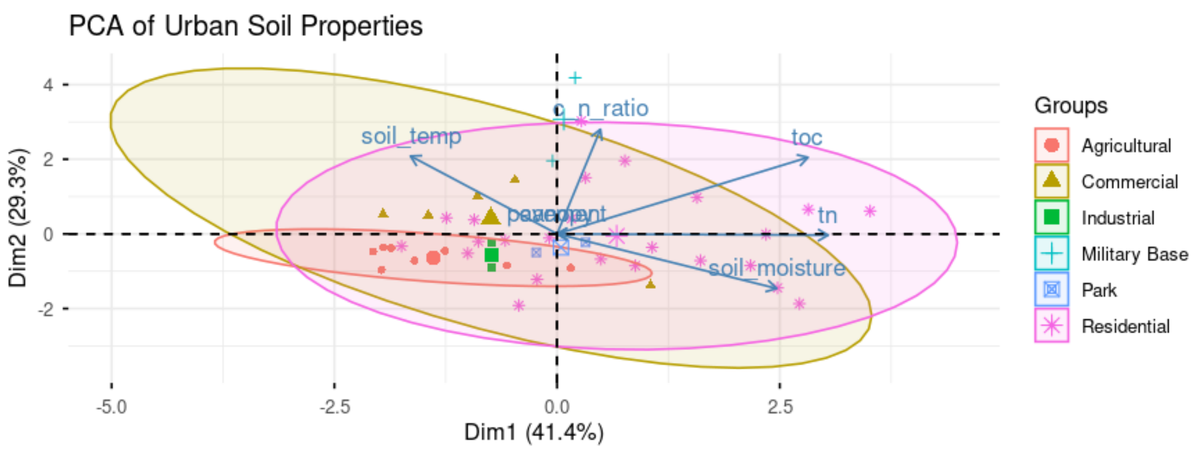

Figure 1. Principal Component Analysis (PCA) for Soil Properties shows two major dimensions (Dim1 & Dim 2) which explain 70.7% of the variance in the data.

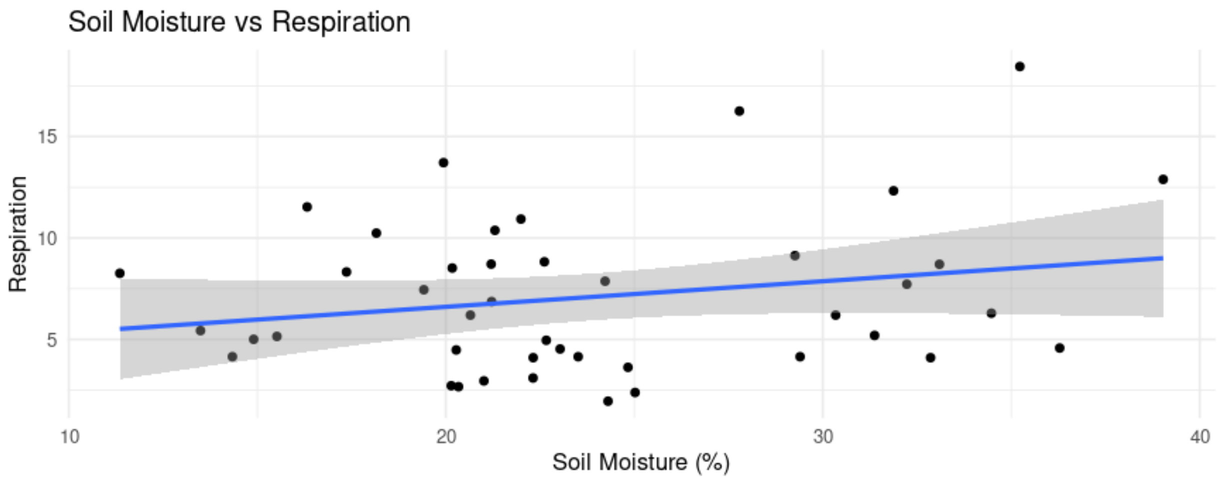

Figure 2. The relationship between Soil Moisture and Respiration shows that increases in soil moisture percentage increases soil respiration across the study site.

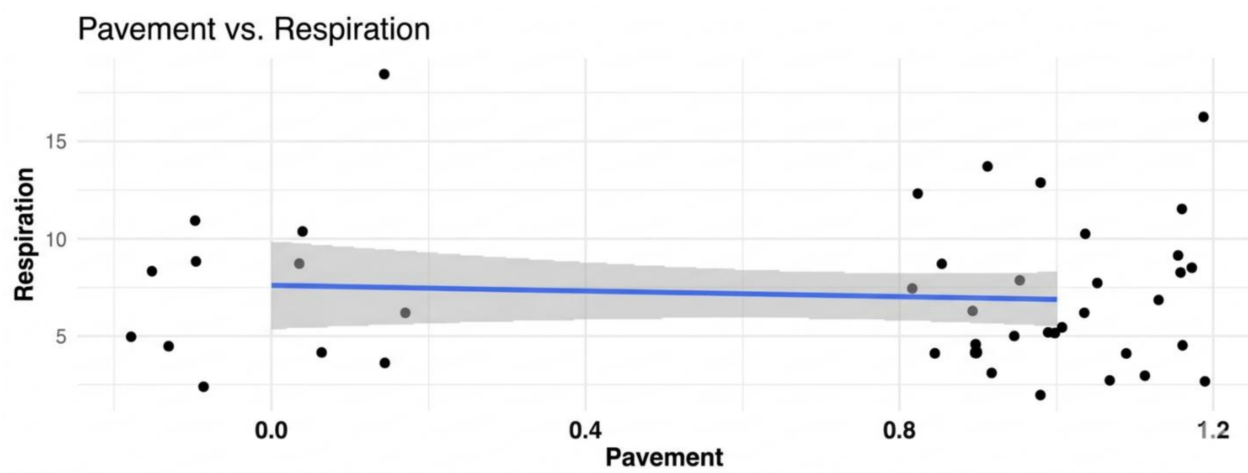

Figure 3. The relationship between Pavement and Respiration shows that pavement drives temperature but is not a major predictor of respiration.

Method

Forty-three soil samples were taken from over twenty unique locations across Davidson, Montgomery, and Robertson Counties to construct a predictive model. The sampling was done at a depth of 10 cm, focusing on the most active microbial region.

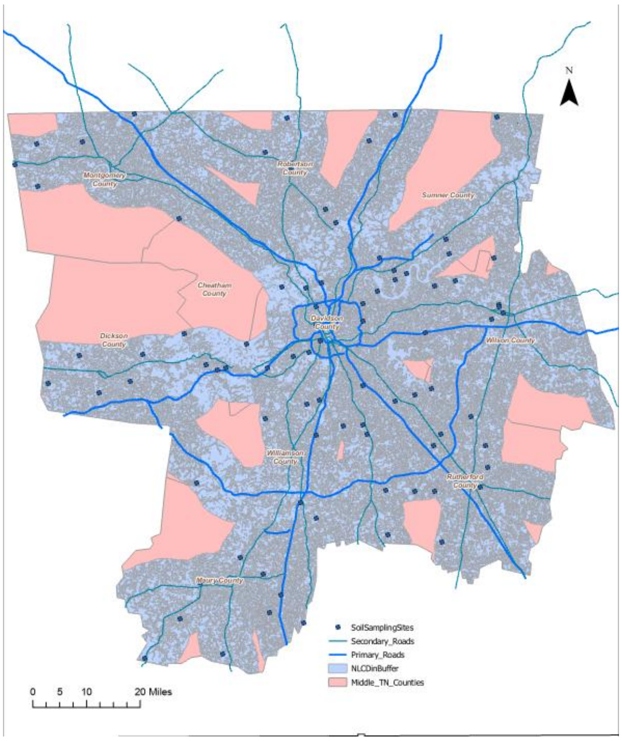

Figure 4. Map of the study site in Middle Tennessee

The feature set includes geophysical variables such as soil temperature, moisture, pH, total nitrogen (TN), and total organic carbon (TOC). Geospatial variables comprise land use (commercial, residential, open space), canopy cover presence, and impervious surface or pavement coverage. Due to the complex, non-linear nature of urban soil chemistry, we applied three algorithms: Random Forest (RF) for handling categorical land use complexity, Extreme Gradient Boosting (XGBoost) to reduce residual errors through iterative processes, and Shapley Additive Explanations (SHAP) for game theory-based assessment of variable contributions. To stabilize modelling with a sample size of n=43, a log transformation was performed on the target variable (respiration), which often exhibits an exponential pattern rather than a linear one. Additionally, 10-fold cross-validation was used to ensure that the model generalized underlying biological principles, rather than memorizing site-specific details from Nashville.

Results

The results of the initial linear models yielded an R^2 of only 0.069, suggesting that urban soil respiration is too stochastic for traditional regression. However, the optimized Random Forest and XGBoost models achieved a Coefficient of Determination (R^2) of 0.9315. This leap in accuracy proves that the drivers of soil respiration are highly coupled; the model needed the depth of tree-based logic to see how moisture and nitrogen interact. Triangulation across all models found soil moisture as the dominant predictor of soil respiration. In **Figure 5**, Partial Dependence Plots (PDP) revealed a critical biological threshold: respiration remains suppressed in dry conditions but shows a massive intensification once soil moisture crosses a 25% threshold.

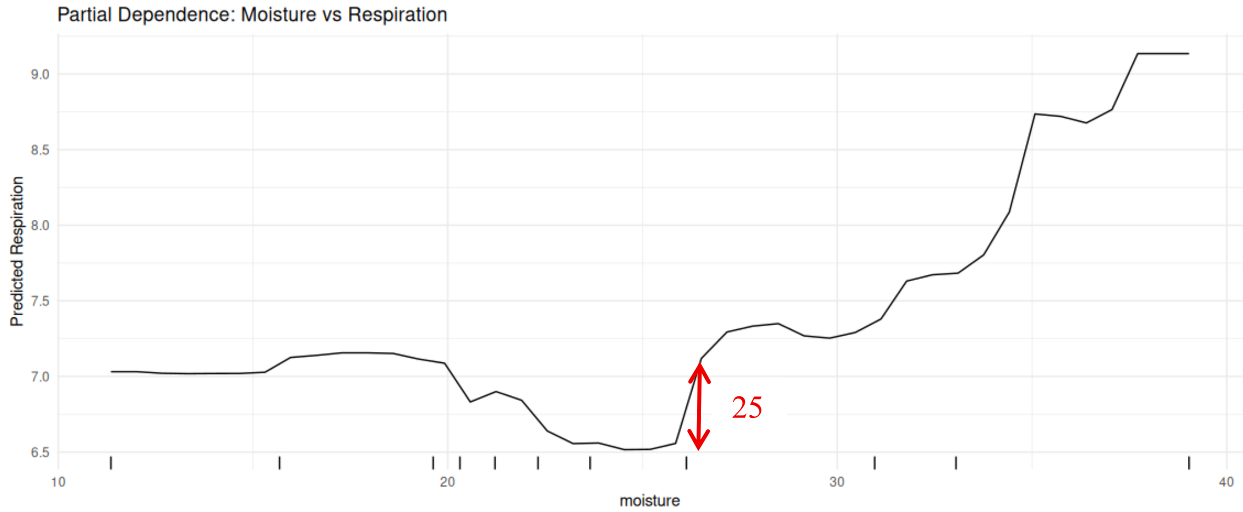

Figure 5. Partial Dependence Plot shows that a threshold of 25% moisture begins to increase exponentially.

This suggests that urban carbon loss in Tennessee may be pulsed and triggered by specific rainfall events that activate dormant microbial communities. Surprisingly, Total Nitrogen (TN) and Land Use consistently outranked Soil Temperature in importance. While the UHI effect increases temperatures, the models suggest that Tennessee's urban soils are moisture-and-nutrient limited, not temperature-limited. In the hierarchy of the urban carbon cycle, what we put into the soil (fertilizers/nitrogen) and how we cover the soil (land use) matters more than the heat of the air above it. **Figure 6** uses Random Forest (RF) model to show variable importance was assessed using two metrics: %IncMSE (Mean Standard Error), which quantifies the loss in model predictive accuracy when a variable is removed, and IncNodePurity, which measures the variable's contribution to reducing data heterogeneity during the model’s decision-making process."

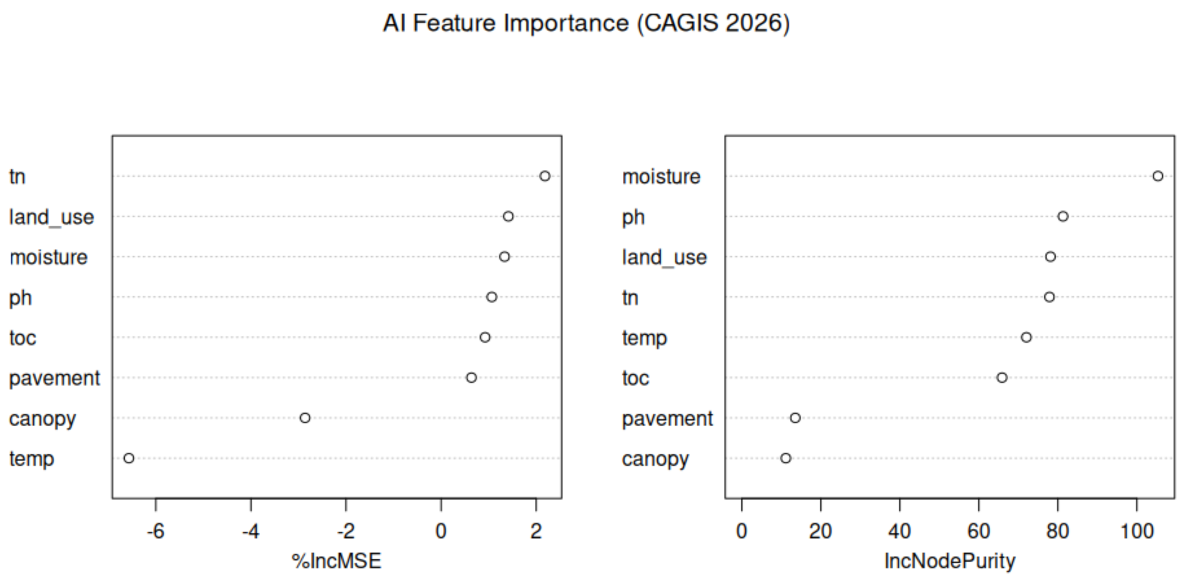

Figure 6. Random Forest (RF) Model shows Variable importance (left) and structural importance

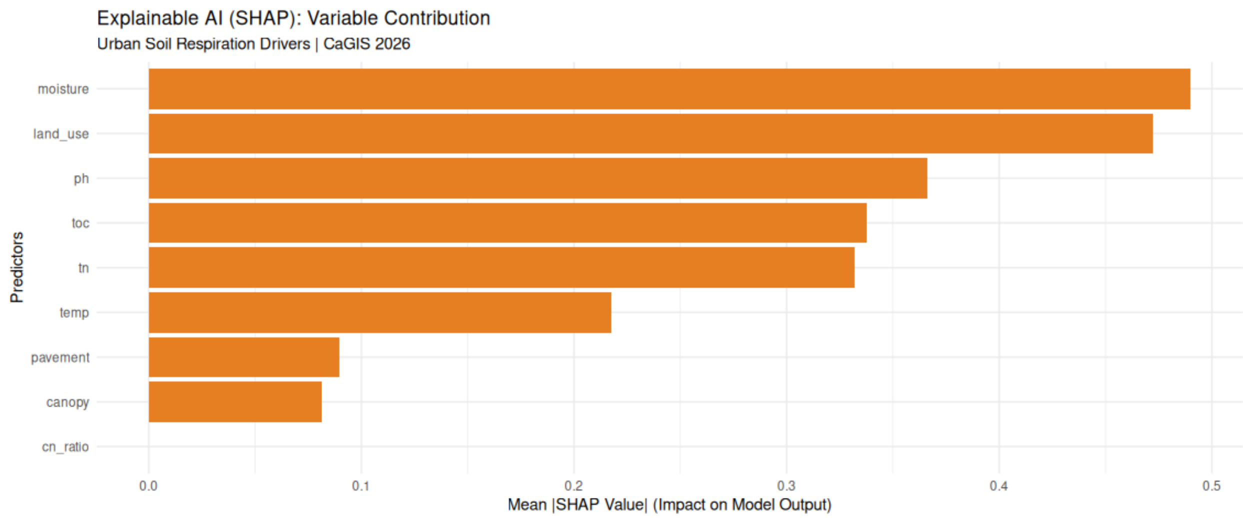

Figure 7. SHAP Variable Contribution plot quantifies the individual contribution and direction of influence of each predictor

In **Figure 7**, SHAP Variable Contribution plot approach quantifies the mean impact of each predictor on the final model output, corroborating that Moisture, Land Use, and pH exert the strongest influence on Nashville's urban soil respiration patterns.

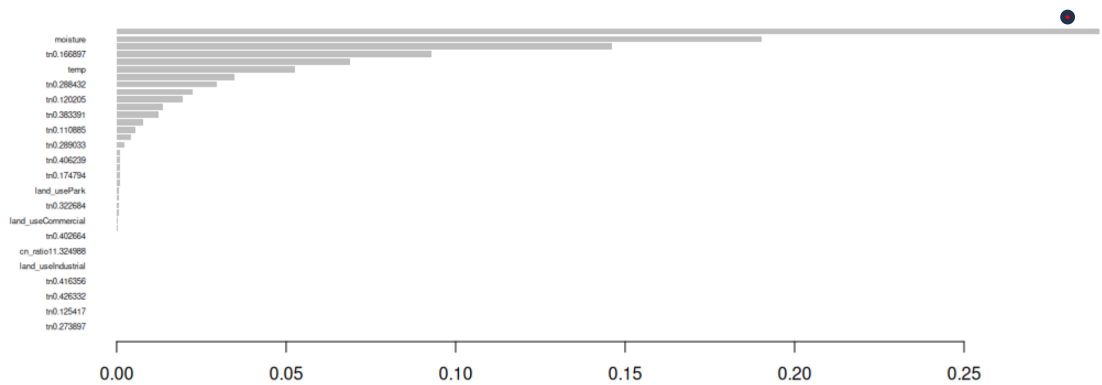

Figure 8. Mean SHAP Values for Environmental Predictors.

Results from the XGBoost model in **Figure 8**, like the SHAP value analysis, identify soil moisture as the dominant predictor of respiration (highlighted with the red dot), significantly outweighing the influence of land use and chemical factors like pH

These findings can help push cartographic boundaries by defining and refining urban and climate policy and even how cartography can be used to communicate research findings. We can visualize hotspots of respiration based on real-time moisture data and zoning maps. For a cartographer, this means the boundary of the map is no longer a physical fence, but a functional threshold (e.g., the 25% moisture line) and for the environmental; agencies such as Tennessee Department of Transportation (TDOT) and urban planners, these results provide a roadmap for mitigation. If nitrogen and moisture are the triggers for carbon loss, then managing urban runoff and fertilizer application in residential zones becomes a working climate mitigation strategy. In conclusion, this study successfully proves that machine learning can decode the complex, non-linear signals of urban soil respiration. With a predictive accuracy of 93%, we confirm that a predictable hierarchy of stressors governs Nashville's carbon flux. By integrating these ML insights into geospatial frameworks, we can better predict how Tennessee's soils will respond to the dual pressures of rapid urbanization and a changing climate. While the current model achieves high internal predictive accuracy ($R^2 = 0.93$), these results are derived from the primary training dataset (n=43). Given the high micro-heterogeneity of urban pedology (human-modified soils), there is a risk of model overfitting. Future research will focus on external validation, by applying this trained algorithm to independent soil temporal sequences, age gradients and development stages across Tennessee to evaluate its generalizability. While this is not a universal predictive tool, the model serves as a robust tool for feature attribution finding the primary environmental drivers within the study area.

Acknowledgements

"This work is supported by the Centres of Excellence at 1890 Institutions program, project award no. 2022-38427-37307, from the U.S. Department of Agriculture's National Institute of Food and Agriculture."

References

Beniston, J. W., Lal, R., & Mercer, K. L. (2018). Assessing and managing soil quality for urban agriculture in ancient and modern cities. *Landscape and Urban Planning*, 170, 25-37.

Crowther, T. W., Todd-Brown, K. E., Rowe, C. W., et al. (2016). Quantifying global soil carbon losses in response to warming. *Nature*, 540(7631), 104-108.

Gasparrini, A., Guo, Y., Hashizume, M., et al. (2015). Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The Lancet*, 386(9991), 369-375.

Grimmond, S. (2007). Urbanization and global environmental change: local effects of urban warming. *Geographical Journal*, 173(1), 83-88.

Harlan, S. L., Chowell, G., Hoffmann, J. P., & Erickson, D. J. (2013). Advection of coolness: Rescaling the social and environmental determinants of health equity in a desert city. *Health & Place*, 24, 233-242.

Hsu, A., Sheriff, G., Chakraborty, T., & Manya, D. (2021). Disproportionate exposure to urban heat island intensity across major US cities. *Nature Communications*, 12(1), 1-11.

Kaye, J. P., McCulley, R. L., & Burke, I. C. (2005). Carbon fluxes, nitrogen cycling, and soil microbial communities in adjacent urban, native, and agricultural ecosystems. *Global Change Biology*, 11(4), 575-587.

Oke, T. R. (1982). The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, 108(455), 1-24.

Treat, C. C., Natali, S. M., Ernakovich, J., et al. (2015). A pan-Arctic synthesis of CH4 and CO2 production from anoxic soil incubations. *Global Change Biology*, 21(7), 2787-2803.

Zhou, L., Zhou, X., Zhang, B., et al. (2021). Different responses of soil respiration and its components to nitrogen addition among biomes: A meta-analysis. *Global Change Biology*, 27(13), 3123-3136.