TriCoMM: Problem 1

Niyati Goswami
Malaika Shroff
Varun Joseph
**Introduction**

UHI– Urban Heat Island– effect is a phenomenon where urban areas, specifically areas with a large amount of concrete, absorb more heat and sunlight as compared to their surrounding areas. Concrete absorbs heat and releases it during nightfall causing the temperature in cities to be greater than their surrounding areas with more tree cover. According to the EPA, “exposure to extreme temperatures increases the risk of illnesses such as heat exhaustion, dehydration, and heat stroke. Older adults are prone to heat-related illness as they are more likely to be in poor health, less mobile, more isolated, and sensitive to high heat. Low-income populations are also at greater risk of heat-related illness and mortality due to lower prevalence of home air conditioning and neighborhood green infrastructure such as urban parks, street trees, and woodlands.”

Durham currently faces UHI and wants to mitigate risks by planting trees in existing green spaces or ones developing. Durham also wants to lend a hand out to low income communities since they have found low income communities are at a disadvantage in regards to UHI. We took into consideration every factor we believed would provide Durham with the most cost effective approach. Since the problem states that working with green spaces equal and/or more than 60% is cost effective as opposed to less than 60%, we determined that there was a correlation between green space and cost. Note that Durham also wanted to prioritize lower income individuals, so after determining the relationship between the amount of green space and temperature, we formulated an equation "Net Benefit = (percent of people below the poverty level + change in green space) - Cost." Since the social benefit is equal to the cost, we determined Net Benefit=0. Then we utilized the model to solve for change in green space, resulting in these final two models:

\[
\text{Net Benefit} = s_1p + s_2 (0.014667)(\Delta X) + s_3c_i(\Delta X)
\]
\[ \Delta X = \frac{s_1 P}{s_3 C - s_2 0.014} \]

Our Model

The first assumption that we made was that because the area of each block was similar, the cost to implement the same percent of greenspace would be the same for all blocks. Furthermore, we assumed that the additional cost needed to add greenspace for blocks with less than 60% greenspace is the same for all blocks regardless of the amount of green space. Therefore, a block with 0% greenspace has the same additional cost as a block with 59% greenspace.

Our first cursory examination of the data given yielded certain relationships between different variables. We began by examining the correlations between each pair of variables in the set and did so with a correlation plot seen in Figure 1.

---

Figure 1
There are clear correlations between the average reduction in ambient nighttime temperatures and the percent of the population below two times the poverty level and the percent of each census block that is greenspace. Further examinations of these variables yielded Figures 2.
These graphs verify a strong linear relationship between all of these variables. It seems clear that there is the strongest positive relationship between the percent of green space and the average reduction in ambient nighttime temperatures as the variance across the data is quite low, but the
relationship between the average reduction in ambient nighttime temperatures and the low income population proportion can’t be easily disregarded. It seems that a linear model can be used to ascertain a strong estimate of what determines the average reduction in ambient nighttime temperatures. To consider all the variables present, a backwards selection process was performed to find the ideal model for a predictor of average reduction in ambient nighttime temperatures.

For the sake of thoroughness, a linear model was formed using every variable available. Using a p value of 0.05 to determine statistical significance, latitude and longitude were easily removed in that order followed by percent of the population below two times the poverty level. The final model is a simple two dimensional linear regression based only on the percent of greenery present. It can be written as the average reduction in ambient nighttime temperature equals negative 0.02468548 plus 0.0146670 times the percent of green space. The p value of this estimation is 2.2e-16, indicating a strong estimate. From this, we can ascertain that a unit increase in percent of green space has an estimated 0.0146670 increase in the average reduction in ambient night time temperature. We use this in our full model later on.

In order to use a linear model, there are certain criteria to ensure that the linear regression procedure we have articulated above is the correct one to pursue. Our first condition is that our data is roughly linear. Our graphs have shown a clear linear relationship between the variables present. The next condition is that the residuals are normal. We can see this pretty clearly from the increased variance on the ends of the lines present in all the graphs in Figure 2.3. Homoscedasticity is another condition and this an assumption we are making as well, as we don’t have a good way of ensuring it is true. Lastly, this model requires independent observations. Each observation applies to a different census block and should act independently.
In reality, it is possible that the reductions in ambient temperature are affected by the greenery in nearby census blocks, and likewise there are relationships between poverty levels in neighboring census blocks and thus social benefit to one block if another neighboring block gets greenery. Because we can’t actually separate a city into geographical and population-wise independent blocks, we are operating under the assumption that each observation is independent for the sake of this model.

For our model, we wanted to create a cost benefit analysis function that takes into account the various conditions we were given. With the conditions given in this problem, we were essentially given certain variables that act as negative and positive influences on our environment. Many environmental models are able to take into account social priorities by incorporating them into a cost benefit analysis, and we were told that the city wants to help low income communities, a benefit we aimed to use as a quantifiable data point. We came up with two primary benefits to consider: the social benefit involved in prioritizing population that tend towards lower incomes and the benefit due to the average reduction in ambient nighttime temperatures. The only cost we consider is the literal cost of implementing such a thing.

For our first social benefit, we decided to use a point system based on the relative percentage of the population under two times the poverty line. Simply put, populations with 100% below the two times the poverty line get 100 points of social benefit. Likewise, populations with 0% below the two times the poverty line get 0 points of social benefit. This allows us to add a benefit to our overall equation that adds to populations that have lower incomes in order to prioritize them.

For the more easily quantifiable benefit of the average reduction in ambient nighttime temperatures, we draw from the linear model created earlier. We know that each percent change
in percent green space has an estimated 0.0146670 change in average reduction in ambient nighttime temperatures, so we characterize the benefit in temperature reduction to be equal to 0.014667 times the percent change in green space.

For our cost, we are given that there is a heightened price of adding green space when a population has less than 60% of its existing area as greenery. We use a constant, $c_1$, to denote cost in our function, and we use $c_1$ and $c_2$ to denote two respective prices. We chose to have two set prices for the increases in percent increase in green space. $C_1$ is the value of the cost of increasing green space by a single percentage in a census block where the existing greenspace is already above 60%. Adversely, $C_2$ is the value of the cost of increasing green space by a single percentage in a census block where the existing greenspace is below 60%.

Finally we add scalars to each cost and benefit present in order to prioritize them in any that is wanted. Because each of our costs and benefits are in different units, with the social benefit being from an arbitrary point system, the other benefit being in terms of a celsius reduction in temperature, and our last variable being a monetary value, we need a way to scale everything to the same respective power. For example, outrageously high costs could make the benefits too negligible to pull any weight, and so we would need a higher scalar for the benefits to equal out the costs. Traditionally, there is a monetary cost designated to the benefits of these models, but we have no good way of quantifying the cost improvement of our social benefit or our temperature benefit, so we leave the model with scalars in place to be adjusted by those who can decide how much each benefit is worth.

This brings us to our final cost benefit model,
Net Benefit = s_1p + s_2(0.014667)(ΔX) + s_3c_i(ΔX)

Our variables in our final model were described as C_i being a fixed, unknown constant for the cost per percent green space. Using the current percent green space, this could be one of two constants: the cost per percent green space for blocks with already existing green space coverage above 60% area, C_2, and the cost per percent green space for blocks with already existing green space covering less than 60% area, C_1. We calculate for X, which is the percent change in green space, or how much green space we should add. P is defined as the percent population income below the poverty line. S_i is the scalar that we are adding to the respective costs and benefits.

In order to use the model, we chose to examine the place where the costs equalled the benefits, or where Net Benefit is 0. Setting this equation equal to zero, we are able to derive:

\[ ΔX = \frac{s_1p}{s_3c_i-s_20.014} \]

This is now an equation where we can simply plug in the assigned values and get a suggested percent green space increase for each census block. It is important to note that these equations are being applied to each census block independently.

In order to see how the model performs we would need to input values for the scalars as well as the costs. These are not values we are presented with or can find for the research. We could use suggested values based on data from the costs of building greenery in other areas but the model would be the most accurate with information about the costs per building greenery in Durham specifically. The scalars are up to be determined by the city of Durham and depends on how much Durham values these benefits. Changes in the scalars can drastically change the model and completely change the change in X. It is not the only influential input however, as holding all else steady, an increase in P increased the change in X which makes sense because as
a census has more people under two times the poverty line, then the more that census block is prioritized for added greenery.

We notice that the strongest correlation we find is between the percent of green space and the average reduction in ambient nighttime temperatures. Our weakest correlations were between low income and average reduction in ambient nighttime temperatures as well as low income and percent of green space. We saw that as P increased, X did as well, which accounts for the government wanting to prioritize lower income communities. Furthermore, we also saw that as C increased, X decreased, accounting for the budget.

As a whole, the weaknesses of our model is due to the oversimplification of many of our variables. There are many assumptions made that boiled down the variables in this situation to a linear or one dimensional model. In reality, most things are more complex. For example, cost is not going to be fixed at a certain specific amount for census blocks above and below sixty percent green space, but we treat it as such in order to simplify our model. In reality, cost would be variable for a course of other reasons and probably also not one dimensional. There could be increases in cost more so as census blocks have less and less existing green space. Our model currently assumes that there is an equal cost per increase in green space percentage for a census block with no existing green space and a census block with fifty nine percent of an existing green space. Likewise, other variables may also need more nuanced real world models. There are also practical limits that constrain our model, like how much can actually be green. Our ideal amount of greenery could come out to a value too high that doesn’t account for the area in each census block needed for housing and business. Furthermore, there could be other variables that influence our model at play that aren’t given in the dataset. Our core strength of this model is the interpretability and flexibility. The simplicity of our model allows for it to be used quite easily
and in quite a versatile fashion. The inputted parameters are few and can be adjusted when moving from block to block. This equation can be used for years to come with separate changes in different parameters. This equation is also broad enough to have applications to more than just Durham. Our use of scalars lets our city determine for itself how much it really cares about certain benefits. Our cost variable can be adjusted throughout the years as the cost of building greenery varies throughout time.

**Conclusion**

There was a lot more information we could have benefitted from. For example, knowing how the cost fluctuates based off of already existing green space along with knowing the exact areas of the blocks could allow the output to be more accurate. However, with the information we do know, we would try to consider changing the model based on if the cost is compounded or if it changes depending on how much the green space changes. As of right now, our cost is one of two unknown constants based off of the information that we were given. If we were able to figure out how the model would change with the cost consistently changing, we would most definitely include that. We are also aware that this model is very simplified in terms of the variables we consider. As we mentioned prior, we know that this data fluctuates drastically along with their relationship with each other in real life, and there are definitely a lot of other variables to consider rather than just assuming the relationship between them is consistent like it is in our model. Our current model attempts to balance cost with benefit, but a higher order equation for net benefit derived from more complex costs and benefits could be used to maximize benefit rather than neutralize costs, which could be more useful to the city of Durham.
Our model tells us that when we determine optimal cost, it correlates with the percent of individuals under the poverty level and change in greenspace combined. We believe that data regarding population per block would help quite significantly. According to Enviro Atlas. “In a city with one million or more people mean ambient air temperatures can be 1.8 to 5.4°F (1 to 3°C) higher than those in the surrounding region.” An increase in consumption alongside population growth leads to an outpour of greenhouse gases. Countries where individuals accumulate the most wealth, who tend to have overwhelming practices in lifestyle and production, generate higher emission rates than countries who are not at the same tier. The US is responsible for 17% of earth’s energy consumption but has only 4% of the world population. According to Wral, starting from April 2020, Durham has been through 14 economic projects and expect 6,900 jobs to bloom from these developments.

References


Dear _____,

Our mathematical model shows the cost-benefit analysis for the percent green space that should be added in each block that shows what we believe is when the social benefit equals the fiscal benefit. We believe that each block has its own unique amount of green space that should be added, and we are very confident that this model shows exactly how much greenspace should be added in order to optimize cost while still making a beneficial impact on the environment.

One of the main factors we need to prioritize is the lower income communities. With this model, we are able to account for how much green space should be added in order to maximize social benefit while also considering the financial cost. With the percent greenspace increasing as percent income below the poverty line increases and percent green space simultaneously decreasing as cost increases, we are able to find the right balance of the two variables to still be able to make a positive economic and social impact without it being too costly. Along with all this, we figured out what the average decrease in temperature would be for every percent green space added.

We believe this model makes the most environmental benefit while considering the cost and the communities that you would like to impact. We look forward to the opportunity to hopefully continue working with you on this project.

Best,

TriCoMM Team 9