TRICOMM PROJECT REPORT

TREEBEARD'S TRIALS: A CONSTRAINED OPTIMIZATION MODEL FOR WHERE TO PLANT TREES IN DURHAM TO MINIMIZE ITS URBAN HEAT ISLAND WITH SPECIAL CONSIDERATION TO UNDERSERVED COMMUNITIES

Enzo Moraes Mescall
Erik Mendes Novak
Nicolás Salazar Awad

Contents

1 Introduction **2**
1.1 Background: . . . 2
1.2 Methods: . . . 2
1.3 Results: . . . 2
1.4 Conclusion: . . . 2

2 Brief overview of our model **3**

3 Assumptions **3**
3.1 Tree growth patterns . . . 3
3.2 Tree logistics . . . 4
3.3 Forests and climate . . . 4
3.4 Selecting which trees to plant . . . 5

4 Building our model **5**
4.1 Utility function . . . 5
4.2 Constructing the components of the utility function . . . 6
4.2.1 Modeling night-time temperature change . . . 6
4.2.2 Defining $\vec{\alpha}$ . . . 6
4.2.3 Calculating tree canopy coverage gains . . . 7
4.2.4 The cost of planting trees . . . 7
4.2.5 The break-in cost . . . 8

5 Parameter estimation for our model **8**
5.1 Estimating the effect of forested area on night-time temperature change . . . 8
5.2 Estimating the parameters for tree crown spread growth over time . . . 9
5.3 Gauging the parameter α . . . 11
5.4 Other estimated variables . . . 11

6 Model deployment **11**
6.1 Greedy algorithm . . . 12

7 Solution and results **12**
7.1 Results . . . 12
7.2 Sensitivity analysis . . . 13
7.3 Strengths and weaknesses . . . 14

8 Conclusion **16**

9 Appendix **18**
9.1 Code . . . 18
9.2 Comments on the EnviroAtlas database . . . 18
9.3 Plotting greenspace that is not forest vs UHI . . . 18

1 Introduction

1.1 Background:

The Urban Heat Island effect (UHI) is a harmful side-effect of urbanization where the prevalence of concrete in cities and deforestation for development lead to average temperature increases due to increased absorption of solar radiation [6]. When this occurs, it may make it more difficult for the region to cool down at night, leading to a cumulative heating effect[6]. This is particularly harmful to the elderly, due to more vulnerable health, and poorer communities, due to less accessibility to air-conditioning and climate control[6]. Focused, goal-driven tree planting has been shown to have a strong mitigating effect on UHI by increasing the overall canopy coverage of a region, hence decreasing absorption[7].

We aim to formulate a recommendation for the Durham City Council on how trees should be planted in the city to meet four goals. **First**, we would like to **reduce the effects of UHI**. **Second**, we would like to **focus our efforts on historically underserved communities**, namely low income communities. **Third**, we want to account for how **tree planting on heavily populated census groups** will have a greater effect on the aggregate population's wellbeing. **Fourth**, we want to formulate our recommendation **constrained by a fixed budget** in order to maximize our impact.

1.2 Methods:

We relied on a utility function U which takes in a price vector $\vec{P}$ having $\vec{P} \in \mathbb{R}^n$ and $U : \mathbb{R}^n \to \mathbb{R}$ where each dimension n of the price vector represents the amount invested into that district. The assumption we made is that, given an accurately constructed utility function, whatever $\vec{P}_{\text{ideal}}$ maximizes U will be the wisest distribution of investments across the Durham area that balances both our ecological and social goals.

To create U, we considered detailed factors like the growth of tree crown spreads, and entry costs to planting saplings, as well as more large-scale factors like the relationship between average temperature change at night and an area's canopy. After creating U, we then built $\vec{P}_{\text{ideal}}$ using a greedy algorithm computational method, since finding $\vec{P}_{\text{ideal}}$ analytically is nonviable due to the verbosity of U.

1.3 Results:

We ran the algorithm for a number of different budgetary constraints to see how budgetary changes would affect the overall positive impact of the initiative, and noticed that at a $500,000 threshold any further increase in budget has diminishing social benefits.

The greedy algorithm constructed results in the following investment recommendations: Census groups 49, 80, 134, 141, 155, and 178 all receive substantial funding which would result in the planting of about 1490 new trees.

1.4 Conclusion:

We propose a budget of at most $500,000 and tree allocations as per our results; however, we acknowledge that actual meaningful UHI mitigation will require long term investments continuously for multiple years, and it will depend on the mayor's priorities and what budget levels are viable.

2 Brief overview of our model

We are working with all 193 census groups in Durham County. Given a pre-established budget and information on the population and income level of each census group, our model outputs a specific allocation of budget to each census group (translating to different numbers of planted trees). This specific allocation **finds a balance between the constraints of a fixed budget, the equity of investing in poorer neighborhoods, the more intense effects that will be felt in a dense census group, and the upshot of increased canopy coverage in general**, via the following:

1. It encodes the effect on UHI that tree planting will have on each specific census group.
2. Accounts for equitable practices by ensuring our budget allocation is weighted heavily by the poverty level of a neighborhood, which translates to improvements in UHI correlating proportionately to the poverty level of a census group.
3. It ensures budget allocation is weighted by the population of a neighborhood, so that more people will feel the benefits of this program.
4. Allows for a flexible exploration of how different budgets will affect the impact from tree-planting.

3 Assumptions

In order to simplify our model, we made the following assumptions.

3.1 Tree growth patterns

G-1. *Crown spread shape*: Our tree crown spread will always be perfectly circular, with the tree trunk centered at the circle's center.

G-2. *Canopy coverage from crown spread*: The canopy coverage for the tree will be uniform and guaranteed underneath the crown spread for the tree.

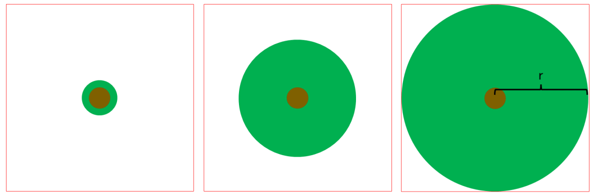
G-3. *Crown spread size*: Each tree species we plant will have the same maximum crown spread size, which is assumed to be reached at the average age when that tree species reaches maturity in real life.

G-4. *Tree growth*: The radius of crown spread of a given tree will grow as a function of time until the tree reaches maturity. Thus, at any point in time, crown spread will be a concentric circle about the tree trunk.

G-5. *Tree size as a sapling:* When a sapling is originally planted, we assume it has a crown spread radius of 0. In reality, it may have a small radius already, and its radius shortly afterward becomes predictable by our model used. Hence, we assume it starts at 0 for better consonance with the model.

G-6. *Tree density*: Trees need physical space in which to grow. We assume the tree needs enough space such that its crown spread will never overlap with the crown spread of another tree. That is, we get a circle packing situation for maximum density.

Figure 1: From left to right, the growth of a tree's crown spread over time



3.2 Tree logistics

L-1. *Hydrologic and nutritional needs*: We assume that any plot of land can always be adequately prepared to house a tree with enough water and nutrients in the soil.

L-2. *Tree planting locations:* In fact, we assume that any area not explicitly considered "forested" as per the official USDA database can equally well accommodate a new tree [5]. At a glance, this would thus also include buildings and bodies of water, a model simplification we had to accept.

L-3. *The planting itself:* Trees are placed instantly. This way, we can simplify the model by not needing to model the logistical components behind the act of planting a tree.

L-4. *Supply chains*: We assume there will be no supply chain shortages for saplings or labor shortages that may delay the planting of trees. Thus, when the program commits to a time frame, it will be invariably reached.

L-5. *Bulk sapling purchases*: It is possible that economies of scale could make the bulk contract purchase of saplings cheaper. However, we simplified the model via a linear uptick in prices after the fixed costs of an initial contract.

L-6. *One-time initiative*: The way our model is currently structured, all investments will happen in a one-time batch, perhaps over the span of a year, rather than a gradated process along many years.

3.3 Forests and climate

F-1. *Mortality rate of trees*: In order to isolate the specific results of tree planting, we chose to set the current mortality and birth rate of forests to 0. Thus, for all intents and purposes, forests in any census group of Durham county will stay static besides any tree planting initiative. That is, there will be no logging, nor any tree deaths or births due to natural causes.

F-2. *Impact of big trees*: A study on the tree planting initiative in Bristol, aiming to plant $250,000$ trees by 2030, shows that planting fewer bigger trees has an overwhelmingly better impact in UHI than planting many smaller trees [1]. As such, we made the conscious choice to include in our plans only trees that are considered at least moderately large.

F-3. *The 10-20-30 rule*: This is a common rule in arboriculture that states that, when planting on a large ecosystem-wide scale, one ought to plant a varied selection of species to avoid vulnerability to pests, soil depletion, and other catastrophic destabilizing forces. In general, the rule of thumb is that the tree variety should include at most 10% from each species, 20% from each genus, or 30% from each family [4]. We will abide by this principle when selecting tree saplings to plant.

3.4 Selecting which trees to plant

In selecting which trees to plant, we must recall that we have a robust dataset with more than $14,000$ trees, of which many are native to North Carolina [5]. In selecting which trees to pick, we want to find trees that are some mixture of strong, long-lived, healthy, and having a big crown spread. For example, the yellow poplar would not be a good choice because it is susceptible to pests and disease, and are considered "weak wooded". Additionally, our tree selections ought to at the very least belong to different genuses, so we need not select too many trees to still respect the 10-20-30 rule.

Upon examination, these are the five trees selected:

- Red maple
- River birch
- Sweetgum
- Southern Magnolia
- Bradford pear

These also have the added benefit of being generally aesthetically appealing, adding to the beautification of whichever neighborhood they may be added to.

4 Building our model

4.1 Utility function

In order to build our model, we decided to create a utility function $U(\vec{P})$ having $\vec{P} \in \mathbb{R}^n$ and $U : \mathbb{R}^n \to \mathbb{R}$. It takes in the following variable $\vec{P}$ where each entry p_i represents the investment in reforestation the ith district and n represents the total amount of districts considered. This utility function acts to estimate the social benefit of our investments.

Since the goal is to optimize budget allocation, we then define a vector $\vec{P}_{\text{ideal}}$ as the value which maximizes the following utility function $U(\vec{P})$:

$$\vec{P}_{\text{ideal}} =: \arg\max_{\vec{P} \in \mathbb{R}^n} U(\vec{P}) \tag{4.1}$$

Where n represents the amount of districts. The challenge is then building a function U. We took a top down approach, looking at the consequences we wanted to model and creating functions for their causes. Thus, we defined the utility function as

$$U(\vec{P}) = K(\vec{P}) \cdot (\vec{I} \circ \vec{\eta}) \tag{4.2}$$

Where I is a vector of length n where its entries I_i represent the percent of the population in a district having yearly income less than two times the poverty line, η_i represents the literal number of individuals in a district having a yearly income less than two times the poverty line, and $K : \mathbb{R}^n \to \mathbb{R}^n$ is a function which takes in investment and returns an estimation of how much that would change the night-time reduction in temperature[2]. The dot product was chosen since it will scale the temperature impact by the local double-poverty rate, matching our priority to plant in historically marginalized communities.

Notably, we are taking the *Hadamard product* of $\vec{I}$ and $\vec{\eta}$, represented by the $\circ$ operator, which is the pairwise multiplication of their components. This is done since each entry contains unique information about a district and we would like to preserve that.

Clearly, U is maximized when $K(\vec{P})$ and $(\vec{I} \circ \vec{\eta})$ are linearly dependent; that is, when there exists some $c \in \mathbb{R}$ such that $c\vec{K}_{\text{ideal}} = (\vec{I} \circ \vec{\eta})$. Of course, it is doubtful one can find exact values for c and $\vec{P}$ such that this works, especially considering $K(\vec{P})$ is unlikely to be a linear function, but we can find the best approximate values for these purposes.

4.2 Constructing the components of the utility function

4.2.1 Modeling night-time temperature change

Our approach to constructing an estimate of K took the approach of creating an estimate for the impact of the number of trees planted on the nightly reduction in temperature. Running a linear regression model with temperature as the response and percent of forested area as the predictor gives us B_1. The parameter predicts how much a one-unit increase in forested area percentage impacts nightly temperature reduction. We do this because we are strictly interested in how much our model changes the nightly temperature reduction. As such, the intersect parameter B_0 is not directly relevant. We may then construct the following function:

$$K(p_i) = \beta_1 \left[r(t)^2 \pi T(\vec{P}) \circ \frac{1}{\vec{Area}} \right]_{\leq \alpha_i} \tag{4.3}$$

Where $r(t)$ represents the average radius of the trees being planted as a function of time, $Area_i$ is the square-mileage of the ith district, and $T : \mathbb{R}^n \to \mathbb{R}^n$ is a function which outputs an estimated quantity of trees planted given a monetary value invested into each district. Once again, $\circ$ identifies the Hadamard product between two vectors. One should note the parameter α_i, which represents the maximum possible increase of forested area percentage of any given area, since we cannot have over 100% of a region being forested and the initial values vary amongst districts. If the value $\frac{r(t)^2 \pi T(p_i)}{Area_i}$ is greater than α, we substitute α into the equation, giving us the following:

$$K(p_i) = \beta_1 \alpha_i \tag{4.4}$$

4.2.2 Defining $\vec{\alpha}$

There is an upper limit to the amount of land that can be covered by the canopy of trees in a district. Taking a broad look at the data reveals a maximum of around 90% to be the upper limit, which is further investigated in section (5.3). α can then be roughly modeled by the non-negative quantity.

$$\alpha_i = [0.9 - \text{Current Forested \%}]_+ \tag{4.5}$$

Such that an increase of $\leq \alpha$ percent in forested area will never lead to a district surpassing the upper bound of 90%.

4.2.3 Calculating tree canopy coverage gains

Our model for tree planting will, given a certain number of new trees n and a number of years t, output the increase in canopy coverage. We built this model so that it can be run individually for each census group, yielding the net increase in canopy coverage for each such group independently of others. This way, we can abstract this model into just an input-output function when building the overall model for tree allocation given a particular budget. Similarly, we can run the model for any time frame, allowing a better understanding of how a time-variance model might affect coverage in the short and medium-term.

We assumed that, in order to grow, a tree could not be planted close enough to another tree such that their crown spreads overlapped. Thus, for our model we will plant our trees such that these circles tessellate as best as possible a bigger ambient area in which the trees will be planted. This ought to maximize tree density and therefore our ability to maximally cover the ambient area.

Now, the literature indicates that the following relationship is key to how crown spread radius varies with tree age[3]:

$$h(t) = a(\ln(t+1))^b \tag{4.6}$$

Here, a, b are parameters to be estimated, while t is the elapsed time.

This function, however, only defines the relationship between age and dbh, which is a measurement of how wide a tree is at the widest point of its trunk, and also defines the relationship between dbh and crown spread radius. That is, our actual relationship will be given by the following:

$$h_1(h_2(t)) \tag{4.7}$$

where h_2 is the relationship between t and dbh and h_1 is the relationship between dbh and crown spread radius.

Of course, these will entail specific values a_1, b_1, a_2, b_2, and so we will need to apply regression twice to later calculate these values.

Now, we still need to consider how to allocate the trees that will be planted. Because of the 10-20-30 rule, and since we only have 5 tree types at our disposal, we must plant rather close to the 20% threshold for each one.

So, given n trees allocated to a district, we intend to plant as close as possible to $\frac{1}{5}$ of them from each of the five species.

However, since the $\frac{1}{5}$ number is not a hard threshold, but rather a guideline, we can simply assume that in any situation we will always be able to approximate this value, or that there are too few trees for any offset to make a difference.

In terms of calculations, however, it is important to note that each tree species will have its own parameter estimates, and so will yield its own $r(t)$ function. Therefore, for $1, 2, 3, 4, 5$ some indexing of the tree species, we ultimately have the following function for $r(t)$ the cumulative radius function:

$$r(t) = \frac{1}{5}\left(r_1(t) + r_2(t) + r_3(t) + r_4(t) + r_t(5)\right) \tag{4.8}$$

4.2.4 The cost of planting trees

The main aspect in constructing a forestation model is defining the cost of planting trees. We settled on a fixed plus variable cost model for cost of trees, thus benefiting larger concentration of trees within a district.

$$T(\vec{P}) = \frac{\vec{P}}{\text{Variable cost per tree}} - \frac{\text{Fixed costs} + \vec{B}}{\text{Variable cost per tree}} \tag{4.9}$$

4.2.5 The break-in cost

We defined the break-in costs, $\vec{B}$ as the cost of the unique additional factors associated with planting trees in impervious land. The proposal indicates 60% greenery as an adequate cutoff for when investing in forestation does not require additional break-in costs.Thus we constructed the strictly non-negative function:

$$\vec{B} = \left[C_{\max} - \frac{C_{\max} * \vec{g}}{0.6}\right]_+ \tag{4.10}$$

Having $C_{\max}$ represent the maximum possible cost of breaking in to a region that has no greenery. Thus we conclude our process of model construction and may move to estimate values for the various constants used.

5 Parameter estimation for our model

5.1 Estimating the effect of forested area on night-time temperature change

Observing the relationships that average temperature reduction at night had with other variables, the correlation with the proportion of an region's area covered by forest is immediately apparent.

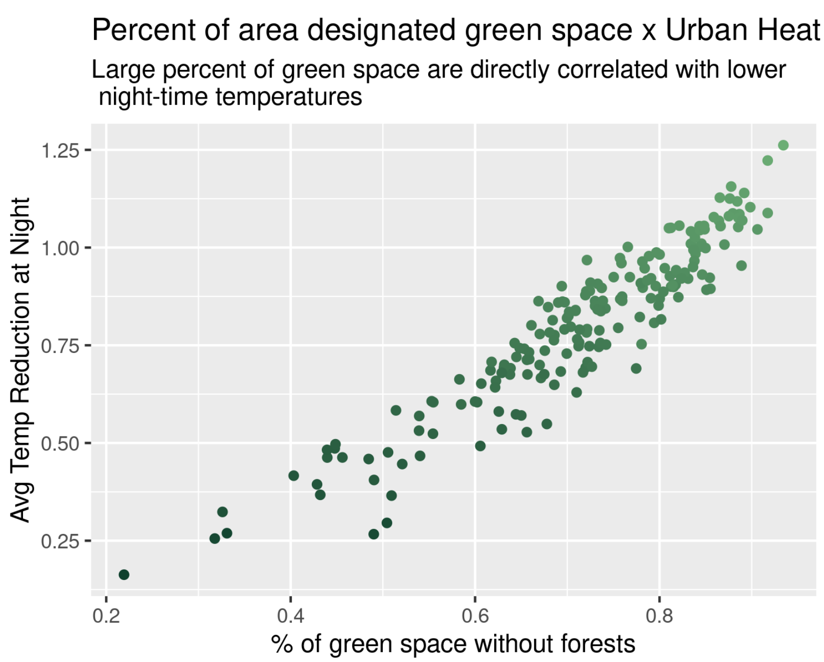


Figure 2: Green space vs UHI scatterplot

More specifically, given the percent of forested area of a region, we can regress the following equation on average night-time temperature change[2]:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon, \quad \epsilon \sim N(0, \sigma^2) \tag{5.1}$$

Resulting in the following:

Table 1: Regression information for avg. change in temperature

	Estimate	Standard Error	t-value	p-value	R^2
β_0	0.0151969	0.0053453	2.843	0.00495	0.9918
β_1	0.0138100	0.0000909	151.929	2e-16	

Thus one may conclude that the relationship is certainly linear enough for β_1 to be used in the utility function.

5.2 Estimating the parameters for tree crown spread growth over time

Recall that we assumed the relationship between age and dbh to be given by $h(t) = a(\ln(x+1)^b$, and similarly the relationship between dbh and crown spread radius to be given by the same formula. Thus, applying this relationship in a composite fashion yields the relationship between age and crown spread radius.

We have a very robust dataset with over $14,000$ trees[5]. Of those, five were identified as being native to the North Carolina region. So, we can run a linear regression on the data set to find estimate parameters for the relationship. For that, however, we must first linearize the model.

For that, recall our original relationship, where $E[h(t_i)]$, the expected value at age t_i, is the dbh at that point.

$$E[h(t_i)] = a(\ln(t_i + 1))^b$$

To linearize this so regression can be applied, consider the following:

$$\begin{aligned} h(t_i) = a(\ln(t_i+1))^b &\Leftrightarrow h(t_i) = ae^{\ln(\ln(t_i+1))^b} \\ &\Leftrightarrow h(t_i) = ae^{b\ln(\ln(t_i+1))} \\ &\Leftrightarrow \ln(h(t_i)) = \ln(a) + \ln(e^{b\ln(\ln(t_i+1))}) \\ &\Leftrightarrow \ln(h(t_i) = \ln(a) + b\ln(\ln(t_i+1)) \end{aligned}$$

So, in the end, setting $\ln(h(t_i)) = Y_i$, and $v_i = \ln(\ln(t_i + 1))$, and finally letting $\ln(a) = A$, we get a linear regression model $Y_i = A + bv_i + \varepsilon_i$.

If we then run a simple linear regression on these values, we get estimated values for $\hat{b}$ and $\hat{A}$, with MSE the mean squared error.

Here, now, we end up with two functions: h_2 takes in time as a variable, and outputs dbh, and h_1 takes in dbh and outputs crown spread diameter. That is precisely what we want, as then $h_1 \circ h_2$ will give the crown spread diameter.

So, we get estimated values for $\hat{b}$, $\hat{A}$, and MSE, both for h_1 and h_2, as in the following table:

Applying our MSE values into our function as in the literature, we end up with the following values, for each tree type (noting that we divide by 2 since we want the radius and not the diameter of

Table 2: The regression parameters estimated for each of the five tree types

	h_2			h_1		
	A_2	b_2	MSE_2	A_1	b_1	MSE_1
Red Maple	-0.365	3.502	0.21332	-0.205	1.995	0.03091
River Birch	-2.717	5.015	0.10769	-0.471	1.989	0.03490
Sweetgum	-0.720	3.509	0.07674	-0.256	2.606	0.03890
Southern Magnolia	-0.010	2.609	0.09677	-0.609	2.095	0.05918
Bradford pair	1.396	1.719	0.12057	-0.782	2.583	0.02399

the crown spread):

$$r(t) = \frac{h_1(h_2(t))}{2} = \frac{1}{2}e^{\frac{MSE_1}{2}+A_1}\left(\ln(\ln((e^{\frac{MSE_2}{2}+A_2}(\ln(\ln(t+1))^{b_2})+1)\right)^{b_1} \tag{5.2}$$

We can also appreciate the functions for the crown spread radii in graph format via the following graph:

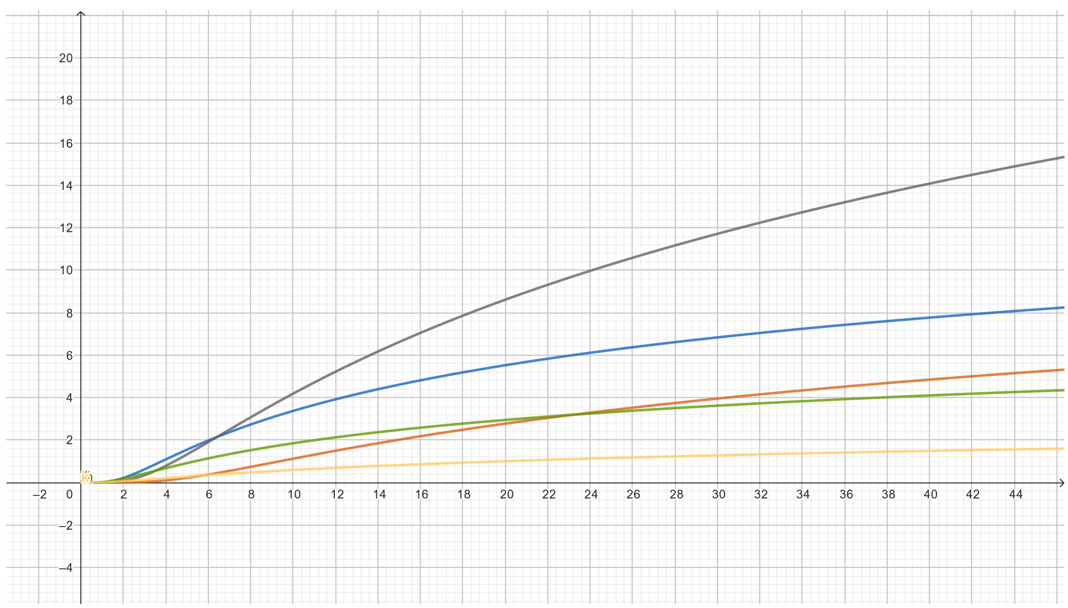


Figure 3: Crown spread radius (m) as a function of time (years) for each tree type (blue is maple, grey is sweetgum, orange is birch, magnolia is green, and the pear in yellow).

Now, the equation above is not bounded above. So, theoretically, a tree that lives indefinitely would also attain an indefinitely large crown spread.

To bound the function above, we chose to manually add a horizontal asymptote at the value of t equal to the tree age at maturity. Thus, whenever a tree reached maturity, its crown spread radius would stop growing.

We chose to deterministically establish this for all trees, instead of creating some sort of Gaussian distribution for tree radii, as ultimately the trees would, on average, end up with this radius anyway.

We thought starting the asymptote at the tree's maturity age ideal because it strikes a delicate balance between allowing the tree room to grow to its full potential, while also capping off its growth at some point. The fact is $r(t)$ exhibits vaguely logarithmic behavior, so by the time the tree reaches maturity in real life (and has thus generally stopped growing) its model also exhibits very little continued growth.

Now, the average age at maturity for the tree species, in order from top to bottom as per the table, is $15, 20, 30, 20, 10$[5]. Letting t equal these values, and then plugging in for the relevant r functions yielded our radii. For example, for the River Birch tree, we got 3.95 meters. As a ballpark, this value also does seem reasonable off the top of our head.

5.3 Gauging the parameter α

We are assuming that trees cannot have their crown spread overlap. The naive approach to doing this is to plant the trees in a grid pattern where they are equidistant. Then, each tree has an enveloping square that does not overlap the squares of any other trees. Then, for a given area, this packing uses up approximately $\frac{\pi r^2}{4r^2} = \frac{\pi}{4} \approx 80\%$ of the space available in the grid.

Going further, assuming a larger budget, we would want to constrain the trees to a smaller area and thus we are faced with the circle packing problem. The optimal pattern for packing circles is known to be in a tessellation of hexagons, the 'honeycomb' pattern. In this layout, the area of the hexagon is given by $r^2 2\sqrt{3}$, so the proportion of the plane covered by the tree's crown spread would be similar to $\frac{pi}{2\sqrt{3}} \approx 90\%$. This justifies the constant 0.9 in equation (4.2.2) for parameter α.

However, this is a simple approximation of reality. The packing problem is dramatically complicated by the fact that we have five tree species, each of which with a different crown spread radius. Packing optimally given these circumstances is a constrained optimization problem in its own right. As an approximation, we feel the 90% value sets a reasonable upper limit to our circle packing context.

5.4 Other estimated variables

The weight for social benefit $\vec{I} \circ \vec{\eta}$ was chosen in part due to the simplicity of the resulting dot product with K and, experimentally, it pushes the recommendation in the right direction without an over-bias to low-income areas that aren't suited or in need for more trees. The estimated one-time break-in cost for a region, C_{max}, and the actual price of trees given by T were built off of an estimation of costs of planting which were gained from a short inquiry with a representative from Kiefer Landscaping Inc.

6 Model deployment

The goal then was to take the theoretical utility function $U(\vec{P})$ and calculate $\vec{P}_{\text{ideal}} =: \arg\max U(\vec{P})$. The first general attempt was to construct U in a manner which it remained differentiable in a box $B \subset \mathbb{R}^n$ so that a global maximum, if it exists, could be found through differentiation. This proved unsuccessful due to the heavy use of ramp functions, multiple modifiable parameters, and the use of the Hadamard product in the temperature change model. An attempt was also made to use gradient descent, but it also failed due to the price of an initial tree planting being over \$200 and the entries gradient of $U(\vec{P})$ being 0 for any entry $p_i < 200$.

6.1 Greedy algorithm

The method we decided to settle on was to implement a greedy algorithm to find a $\vec{P}$ which was able to maximize U. The goal was to allocate small quantities to whichever district caused the greatest increase in utility, and repeat this process until the budget was depleted. Below is some pseudo-code which summarizes the process.

Require: $s \geq 0$ ▷ s will be our step size, our budget is divisible by s
Require: $U(\vec{P}) : \mathbb{R}^n \to \mathbb{R}$ ▷ U is our utility function
  Budget $\leftarrow$ 120000
  $\vec{P} \leftarrow \vec{0}$
  while Budget > 0 **do**
    Max Utility = 0
    Max Index = 0
    for i in $\vec{P}$ **do**
      $\vec{P}_{\text{temporary}} \leftarrow \vec{P}$
      $p_{\text{temporary, i}} \leftarrow p_i + s$
      Temporary Utility $\leftarrow U(\vec{P}_{\text{temporary}})$
      if Temp Utility > Maximum Utility **then**
        Maximum Utility $\leftarrow$ Temporary Utility
        Maximum Index $\leftarrow i$
      end if
    end for
    $p_{\text{max index}} \leftarrow p_{\text{max index}} + s$
    Budget $\leftarrow$ Budget $-s$
  end while
  return $\vec{P}$

The actual implementation includes a clause which limits the total amount of budget per district. This was done out of political consideration, since it would be unwise to dedicate an entire municipal project's budget to a small portion of the city's population.

7 Solution and results

7.1 Results

Given an initial budget, the greedy algorithm is able to present an estimation of the ideal distribution of investments which maximizes utility. Above we have the investment recommendations for the algorithm run at a budget of \$120,000.

We may use the $T(\vec{P})$ function defined in equation (4.2.4) to then estimate that $T(\vec{P}) \cdot \vec{1} =$ 1490 trees would be planet throughout this campaign with the given budget.

An advantage of the approach we took is that we are then able to easily estimate the increase in U as we increase the total budget, seen in the graph below.The derivative of this plot seems to decrease around \$500,000, indicating any investments beyond this point start having more significant diminishing returns. Thus, our recommendation would be to provide a budget no greater than \$500,000

Table 3: The regression parameters estimated for each of the five tree types

Census Block Index	Budget (USD)	Trees planted	% population below income threshold
49	19,000	237	90.43%
80	20,500	254	49.16%
134	20,500	254	50.08%
141	20,500	254	100%
155	20,500	254	45.35%
178	19,000	237	63.98%

and that a level of around $120,000 annual investment could easily be doubled or tripled and the social benefit would increase alongside the budget.

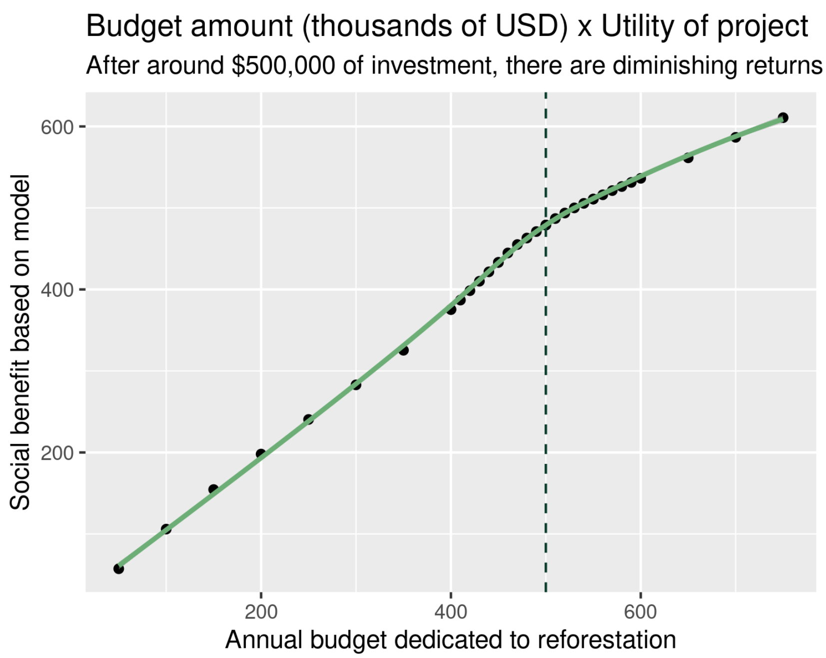


Figure 4: Budget impact comparison for budgets between $50,000 and $1,000,000, in increments of $50,000

The derivative of plot (4) seems to decrease around $500,000, indicating any investments beyond this point start having more significant diminishing returns. Thus, our recommendation would be to provide a budget no greater than $500,000 and that a level of around $120,000 annual investment could easily be doubled or tripled and the social benefit would increase alongside the budget.

7.2 Sensitivity analysis

In sensitivity analysis, we alter values for our independent variables, and see how those affect our dependent variables. In our case, our dependent variable is $\vec{P}_{\text{ideal}}$, which is the argmax of the utility function. However, we want a clear language for comparison when we vary our independent variables.

So, we will instead consider the norm of $\vec{P}$, given by $||\vec{P}||$, as that will symbolize the net mitigation of UHI across the community.

We will conduct a One-At-A-Time (OAT) sensitivity analysis on our independent variables for the model. Specifically, we will conduct tests using $0.60, 0.80, 1, 1.20, 1.40$ as factors for each of the following variables: $\vec{I}$, $C_{\max}$, $\vec{\alpha}$, while keeping the others constant. As mentioned before, our measured value will be $||p||$. Results are below:

Table 4: OAT analysis of independent variables

Independent variable	0.60	0.80	1	1.2	1.4
$\vec{I}$	+0%	+0%	47481.58	+0%	-0.4%
$C_{\max}$	+0%	+0%	47481.58	+0%	+0%
$\vec{\alpha}$	+0.4%	+0.07%	47481.58	+8.6%	+8.6%

Notably, it is difficult to do OAT sensitivity analysis in a multivariate function and the norm of estimated $\vec{P}_{\text{ideal}}$ tends to remain unaltered because of the algorithm's tendency to concentrate the majority of the budget into a few districts, so even if the districts chosen for investment are completely different for $0.6\vec{\alpha}$ and $1.4\vec{\alpha}$, if the budget is distributed in a similar number for both it will seem as if the model is more stable than it is.

7.3 Strengths and weaknesses

Strengths:

1. Our model can be **generalized for a broad range of budget and investment caps** per district, making the model adjustable to any atypical census groups in Durham and even extensible to cities beyond just Durham county.

2. Our model **follows best practices of code modularity**, allowing for easy modification and tweaking of the functions that determine our distribution. That is, we can easily bootstrap new considerations onto existing vertices to add nuance to our model. For example, it was straightforward to modify our model to account for population in each census group once we gained access to that data range.

3. We made effective use of a **greedy algorithm**, which very quickly optimize our budget distribution, thus allowing for more effective exploratory analysis on the effects of budget changes in our outcomes. Since our optimization involves a particular type of monetary distribution, it lends itself particularly well to this algorithm class.

4. Our model makes **deep use of existing large-scale databases**, both the USDA trees data set, with over $14,000$ trees available, and the EnviroAtlas data set for Durham County, from which we extracted all relevant county-specific data like poverty levels, census group area sizes, and more [2] [5]. Having such robust data as the backbone of our paper lends it weight and consistency.

Weaknesses:

1. We modelled at a deep level of detail the crown spread formation for trees, given time t, but ultimately **assumed trees were planted fully grown**; that is, our intricate model was only used to calculate the crown spread of mature trees, even though it has the power to do much more with how they grow over time. A future direction to explore would be **developing a function that measures how utility changes from year to year as trees grow**, and hence form a greater percentage of greenery.

2. In the EnviroAtlas data set, each census group had a percentage of its area that was green but not forest, and a percentage that was impervious [2]. **Accounting for how unforested area in a census group is not all equal** could help optimize tree placement in districts: districts with substantial amount of green unforested areas would be better candidates than those with little amount of it. We did not do that at all. In the future, we ought to append to the cost function to account for these variations, especially since we already have access to them anyway.

3. In our model, we only specify how a budget should be spent on trees in order to maximize our utility in one year of planting. Ideally, we would want to input this data back into our model, and create a new distribution function for the second year to allocate our trees, then input this back into our data, and so on. In the future, **our model would greatly benefit from being able to iterate over successive years**.

4. Our model **did not consider other forms of UHI reduction**, such as parks. In the future, it would be essential to consider the effects that building a park would have vs planting trees, and incorporating this into our model.

5. Our model **did not take into account tree deaths and the necessity of their replacement**. Accounting for this would likely change our budget distribution as the City of Durham replaces trees that die, and should be incorporated into the model in the future.

6. Our model **assumed that the marginal cost of planting additional trees remains constant as quantity increases**, which goes against known microeconomic trends which would otherwise indicate an increase in marginal cost after a certain quantity threshold. The consequences of this can be seen in the simulations with larger budgets, as they don't present significant diminishing returns.

7. We **did not quantify the positive externalities that planting these trees would bring**, including saving energy from cooling, better water filtration, and offsetting carbon emissions. Quantifying this value would provide insight into how much value is truly generated by planting trees, which could provide incentive for further fundraising and investment.

Overview:

Overall, our model excels as determining the ideal distribution of trees given a budget, income levels, greenery level, area, and population for any given inputs. It can be adapted to any data set with the same variables, and can be easily tweaked to include new variables. It also deals well with various different levels of any of these variables.

However, to improve further, we need to better incorporate how all of these variables could change over time, and we need to be able to iterate our model over successive years. It would also need to address other common methods of UHI reduction, as well as their feasibility of being introduced in each district.

Generally, it seems that to the extent that our models used the correct framing structures (such as the greedy algorithm, for example, being an ideal candidate for our purposes), we made some key generalizations that may have inhibited some nuance in our results.

8 Conclusion

Our model outlines subregions in census districts 370630000000 and 371350000000, delimited by specific indices in our data. $20,500 should go to each of indices 80, 134, 141, 155, and $19,000 should go to indices 49 and 178.

In our model, we set a budget of $120,000, which allows for the planting of 1500 trees. However, our analysis reveals that benefits increase greatly with increasing budget up to around $500,000. As such, **we have two recommendations**. Our first is that **trees be planted according to our distribution above**, which prioritizes low income and populous districts. Our second recommendation is that, over the course of the next few years, **we increase the budget for combating UHI up to $500,000**. In addition to reducing the UHI effect and increasing Durham's Tree Equity score, the planting of so many trees would have various positive externalities, including better filtering of storm water, offsetting carbon emissions, and providing substantial energy savings due to the reduction of UHI.

References

[1] *Achieving tree canopy cover targets: A case study of bristol, uk*. https://www.sciencedirect.com/science/article/pii/S161886672100323X. Accessed: 2022-11-13.

[2] *Enviroatlas data set*. https://www.epa.gov/enviroatlas/forms/enviroatlas-data-download. Accessed: 2022-11-13.

[3] *Equations for predicting diameter, height, crown width, and leaf area of san joaquin valley street trees*. Accessed: 2022-11-13.

[4] *Urban design: The 10-20-30 rule for tree diversity*. https://www.smartcitiesdive.com/ex/sustainablecitiescollective/urban-tree-diversity/116641/. Accessed: 2022-11-13.

[5] *Urban tree database*. https://www.fs.usda.gov/rds/archive/catalog/RDS-2016-0005. Accessed: 2022-11-13.

[6] *Average reduction in nighttime ambient temperature*, Washington DC, 2020.

[7] S. B. Kurn, D., *The potential for reducing urban air temperature and energy consumption through vegetative cooling*.

9 Appendix

9.1 Code

Code is viewable here.

9.2 Comments on the EnviroAtlas database

For this competition, we were provided a preliminary filtration of the EnviroAtlas database. However, its census group numbers did not transfer properly to our dataset. Additionally, we felt we would be benefitted from additional data, such as population and area for each of the census groups.

So, we downloaded the full database from the website, and filtered it more coarsely than before to get more options.

For reference, our database can be found in the GitHub linked above, as the "dataset_final.csv" file.

9.3 Plotting greenspace that is not forest vs UHI

One of our theories to make the model more nuanced in the future is that we ought to focus tree planting efforts on green area that is not forested, as the alternatives are impervious surfaces, which likely have far larger overhead costs, or bodies of water, which are obviously impossible.

To test this theory, we graphed the relationship between UHI and green space that is not forest. There is a clear negative correlation on display here, indicating that there is gained UHI mitigation from planting trees even in already green spaces.

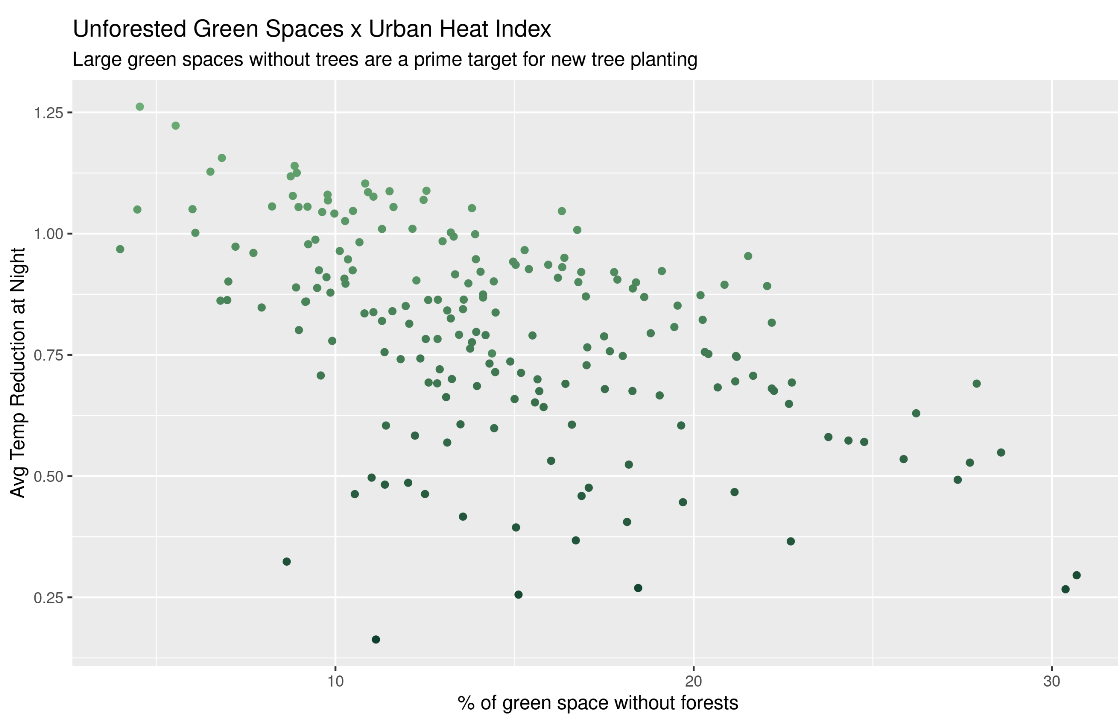


Figure 5: Graph of unforested green spaces vs. UHI