TreesAI

Green Urban Scenarios Framework

Model Specifications Document
Version 1.0
October 28, 2022

*Bulent Ozel and †Marko Petrović

Abstract

This document lays out detailed specification of models, an innovative and practical scenario analysis and impact estimation schemes for green infrastructures. Using the framework, practitioners can design a new urban forest or explore the impact of an existing one. They can investigate each ecosystem service of an urban forest, choosing the level of granularity and complexity on a project-by-project basis.

Keywords: Urban Forest, Nature-based Solutions, Digital Twins, Agent-based Modelling.

*Bulent Ozel, PhD, Lucidminds AI (bulent@lucidminds.ai)
†Marko Petrović, PhD, Lucidminds AI and University of Valencia (marko@lucidminds.ai)

Contents

1 Introduction

Trees and other green infrastructure are a critical part of urban designs. Alongside other infrastructure such as bridges, roads and rails, trees should also enable for investment, profitability and sustainability and should be acknowledged as assets rather than liabilities. To facilitate the transition toward sustainable investments into Nature Based Solutions (NBS), we have developed an innovative and practical scenario analysis and impact estimation schemes called Green Urban Scenarios Framework (GUS).

This document provides the technical specification for the novel Green Urban Scenarios Framework (GUS), which we have developed at TreesAI, to facilitate cities to design, forecast, and monitor green infrastructure portfolios and their long-term impacts under varying weather conditions, maintenance regimes, species compositions, spatial distributions and their exposure to diseases.

Our point of departure is based on the observation that planning, maintaining and estimating the impacts of an urban forest is complex and hence requires a complex systems approach.

Given the level of complexity, we use an agent-based modelling framework in which we combine policy intervention, planning, impact forecasting and monitoring. The entire ecosystems are mimicked digitally, which enables us to capture the context and conduct very granular computational experiments. In short, such modelling features suggest that agent-related abstractions allow us to model a green infrastructure as a complex system with its components and interactions within a geophysical context.

The paper is structured as follows: section (2) provides the description of the core simulation model, section (3) describes the impact analysis, while section (3) concludes.

2 Model

The Green Urban Scenarios framework (GUS) has been developed to facilitate planning and maintenance of urban forests and to estimate their impacts. It consists of a core simulation engine developed following the complex system approach, which is able to simulate the entire ecosystem of trees over N years. It captures the trees' population dynamics, allowing for different case studies and scenario analysis. The framework also incorporates independent modules that use the population data as inputs along with weather data to estimate ecosystem services and benefits. In addition, the framework incorporates a weather data processing module that is used to format local hourly weather data. Figure 1 presents the Trees AI, Green Urban Scenarios (GUS) framework architecture.

The scenario analysis starts from the right-hand side of figure 1, entering the description of a NbS project. The description is given in terms of input parameters to the model describing the size and the characteristics of the population (e.g. number of trees, species, typology, density), planting/maintaining activities, etc. If the typology already exists, the population is initialized reading the data of the population while using the rest of the input parameters given in the NbS description. The initialization procedure is described in details in the following subsection. The initialized tree population, including all parameters along with weather data, is passed to the core simulation engine in $Module$ 1 where the tree population growth and development is simulated over N years. The main component of the simulation engine is a flexible agent-based model that simulates a population of trees at different levels of granularity. If needed, the model can create a digital-twin out of a green urban typology and simulate a spectrum of interactions among the objects (so-called agents) forming that typology. Agents can be trees, humans, invasive species, diseases, etc. All agents are modelled, creating a network among them and taking into account their dynamic interactions also exposed to different weather conditions. Finally,

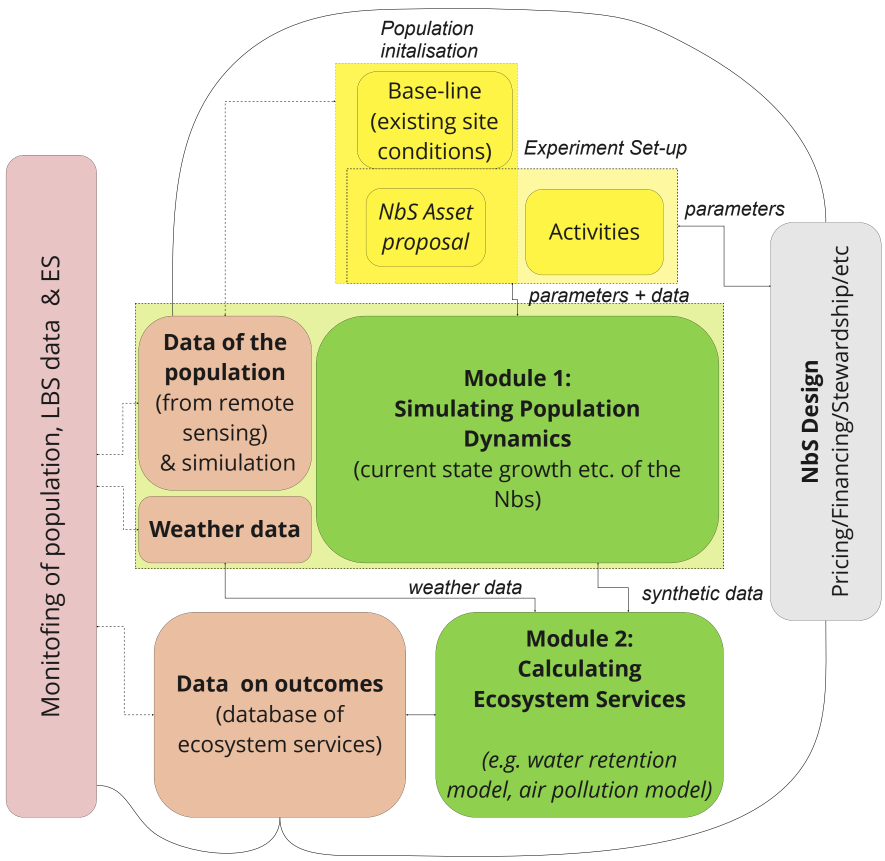

Figure 1: Trees AI, Green Urban Scenarios framework (GUS) architecture diagram.

the tree population dynamics emerge out of interactions among agents in the given typology. Another component of the module is a tree growth model developed following the existing literature (Lorimer et al., 1992; Nowak, 1994, Smith and Shifley, 1984; DeVries, 1987; Nowak, 2020, etc.) which will be explained in details in the following sections. The core module provides synthetic data as output, which are passed to $Module$ 2 along with the hourly weather data from the weather data module to estimate the ecosystem services. Currently, $Module$ 2 includes water retention model whose description is provided in section 3.2 as well as air pollution removal model described in section 3.3. $Module$ 2 calculates ecosystem services and provide data on outcomes to the monitoring activities module, which can be further used to update population data and/or NbS design and to re-simulate the population dynamics and to re-calculate the ecosystem services. Since the nature of the model is stochastic, the simulation can be performed M times over N years, taking into account different weather conditions and different probabilities of agents' interaction. The results are then aggregated over two dimensions, time and simulation runs, and presented in the coherent way. The entire simulation process can be described in four steps: *(i)* inputs; *(ii)* model dynamics; *(iii)* tree growth process; and *(iv)* outputs. In the following, we describe the simulation process.

2.1 Inputs

To make our framework flexible and applicable to any tree typology at any location, we divide the inputs into four different segments: *(i)* site configuration, *(ii)* population configuration, *(iii)* allometric equations, and *(iv)* experiment configuration; where we design an independent (stand alone) database for each segment.

2.1.1 Site configuration

The site configuration is a database with non-exhaustive information about specific sites. For instance, it stores the data about the exact location of a site, its boundaries, the total area of the site, the data about the surface, and any other relevant data that may describe a specific location. The list of variables is expandable therefore, users may use the database to improve the quality of descriptions of the existing sites or enter information about new sites.

2.1.2 Population configuration

The population configuration is a database describing the tree population at a specific location. It can consist of the existing trees at a given location, the planned tree population, or any hypothetical tree population that could be studied at a particular site. The granularity of the data in the database is flexible, meaning that one can describe in detail each tree in the population using a non-exhaustive list of variables such as species, DBH, tree height, canopy height, canopy width, Leaf Area Index (LAI), Bark Area Index (BAI), tree dieback ratio, tree age, canopy overlap, the exact coordinate of each tree in the site, etc. On the other hand, if high granularity is not needed or some data points are missing, one can use less granular data such as more generic species information or, DBH and other tree size measures drawn from a particular distribution. If missing, some of the variables can also be estimated, such as tree height, canopy height, canopy width, Leaf Area Index (LAI), Bark Area Index (BAI), while others such as tree dieback ratio, tree age, canopy overlap, and coordinates can be assumed.

2.1.3 Allometrics

The allometric database consists of allometric equations for specific species at specific locations. Each species at each site is described with allometric equations. If some allometric equations are missing, we define a representative or mean allometric for the specific area, taking into account a more generic family of species. In the most generic case, there is one mean allometric for deciduous and one for conifers at the specific location. This database is extended and updated with newly available data.

2.1.4 Experiment configuration

The experiment configuration database describes the experiments to be performed on the particular tree population at the given location. For the time being, there are available experiments related to the different maintaining strategies. One can choose three levels (M_0, M_1, M_2) of tree maintenance which is related to the planting, removal and disposal, and tree replacement. Maintaining level M_0 disregards any of maintaining activities, while the activities increase with maintaining levels M_1 and M_2.

2.2 Model Dynamics

The core of GUS framework is an agent-based dynamic-stochastic model. The model creates a digital representation of urban ecosystems and using weather data for specific locations it simulates the ecosystem growth and development over time. The digital ecosystem is composed of trees and other participants, so-called "agents", such as people, diseases, invasive species, etc., who are mutually interacting. The model dynamics thus emerge from the agents' interactions given the weather conditions over time. GUS is also a stochastic model. It provides outcomes that are driven by stochastic events or events occurring with probabilities. For instance, weather conditions are stochastic events, while agent actions and interactions are defined involving stochastic elements, i.e., they can happen with certain probabilities. Therefore, the outcomes of the model emerge from complex, stochastic and non-linear agents' interactions for which the analytical solution is not available. Hence, GUS simulates the ecosystems where the entire simulation process involves four steps: (*i*) initialization; (*ii*) spatial configuration; (*iii*) weather projections; and (*iv*) simulation and data collection.

2.2.1 Initialization

Based on inputs (described in section 2.1), the model initializes the digital ecosystem, creating a digital representation of each tree in the population. Each tree is described by a list of variables (DBH, tree height, canopy height, canopy width, ...) from the population configuration database and each tree is assigned an allometric equation from the allometric database, based on species and location. The digital trees are then distributed on a digital grid.

2.2.2 Spatial Configuration

The model uses a spacial configuration method which, based on the site configuration database, creates a digital site. It allocates each digital tree on a grid, where each tree has its own x and y coordinate. Therefore, each tree has a unique place in the digital site, and it is surrounded with neighbouring trees and other agents. The spatial configuration translates all other site characteristics from the site configuration database into digital space, such as soil type, site size, the distance among trees and, the sun exposure.

2.2.3 Weather Projections

The weather data processing module enables weather projections. It uses past hourly weather data for a specific location as inputs and provides hourly weather predictions for N future years, which are used to simulate tree population growth and to calculate ecosystem services. Since the tree growth is simulated on a yearly basis, the weather processing module provides only the information about the frost free days to the simulation engine. Note that, in the case of tree growth, many variables such as the sun radiation, amount of rain, etc., are captured as averages by location specific allometric equations, while some of them are, in fact, taken into account at the micro (tree) level as a growth adjustment factors, like the sun exposure. The rest of the weather data, such as start and end of leaf-on-leaf-off seasons, temperature, atmospheric pressure, relative humidity, wind speed, dew point temperature, the sun radiation, air pollution concentration, mixing height and, atmospheric sounding data (atmospheric profiles for specific locations) are provided on hourly basis and used to calculate the ecosystem services described in section 3. The weather predictions allow us to study how different weather conditions affect the capacity of trees to provide ecosystem services. For instance, one can examine what would be the capacity of trees to reduce run-off during more severe rain events that are expected in the future; or one can examine how the increase in air pollutants affect tree growth and at the same time the capacity of trees to remove pollutants from the air. Although, these effects are non-linear, the simulation of different weather condition along with the simulation of tree growth by GUS allow us to capture those non-linearities for any urban tree typology.

2.2.4 Simulation and data collection

The model is simulated on a yearly basis. Each simulation consists of N iterations (steps), where each iteration represents a period of one year. Within each iteration, each agent is performing tasks imitating what a real agent (e.g. a tree) would in fact do during a year. For instance, trees would be exposed to the weather conditions that would trigger their growth; they would be competing with other trees for the sun exposure which would reduce their growth; they would be exposed to various diseases which would affect their diebacks and probability to die; if a tree dies, humans would proceed with site maintaining actions which could result in removing the dead tree, replacing it or can result in no action, etc.

One simulation of GUS over N years provides one unique realization of the world. Nevertheless, repeating the simulation process, one can obtain another possible/unique outcome. Simulating the model M times over N years, GUS provides the outputs over N years and M unique possible realizations. Hence, the outputs can be analysed either observing each unique realization, so-called "run", "simulation" or "seed"; or aggregating the outputs and looking at the average outcome and its variance over M seeds, which allows us to measure the confidence interval and the risk of the estimation.

In the following, we describe the tree growth process that occurs within each iteration/simulation step, i.e. on a yearly basis.

2.3 Tree Growth Process

The tree growth model predicts DBH, *height*, *crown width* and *crown height* growths taking into account environmental and tree health conditions. It also calculates tree biomass as well as Net Carbon Sequestration (NCS henceforth). NCS is a difference between Gross Carbon Sequestration (GCS henceforth) and decomposition, where GCS represents the amount of sequestrated carbon dioxide in one year, while decomposition denotes the emission/release of carbon dioxide due to tree decomposition process. In the following, we present each segment of the tree growth process.

2.3.1 Initial state

At the initial state, the entire digital ecosystem is initialized, and the system is ready to be simulated. The initial state is denoted as time $t = 0$. Therefore, at the time $t = 0$ all agents' variables are set to their initial values. When the simulation runs from time $t = 0$ to time $t = 1$ i.e., one iteration takes place, agents perform the tasks explained in the following.

2.3.2 Monitoring weather and growth season

Each tree reads info about weather conditions for the current year. In particular, each tree gets the information about frost free days ffd_t at year t.

2.3.3 Monitoring exposure to light and other resources

Trees compute their Crown Light Exposure (CLE) in a dynamic manner. Each tree at each iteration checks the state of its neighbouring trees and determines its current cle_t. Trees first compute the crown overlap at time t with respect to their neighbours:

$$overlap_{ij} = max[0, 0.5 \cdot (crown\ width_i + crown\ width_j - bark\ distance_{ij})][m], \tag{1}$$

where index j denotes a neighbouring tree while index i stands for the reference tree. Note that index t is omitted for the sake of better readability. Therefore, $bark\ distance_{ij}$ is a distance between the tree i and its neighbour tree, j while $overlap_{ij}$ is the crown overlap in meters between trees i and j rescaled for a multiplier of 0.5, to correct for the square shaped assumption of tree crown. Based on the crown overlap size, the overlap ratio is calculated as:

$$overlap\ ratio_{ij} = 0.25 \cdot min[1, \frac{overlap_{ij}}{crown\ width_i}], \tag{2}$$

where 0.25 multiplier accounts for one of the four sides of the grid cell. In addition, the crown overlap ratio is adjusted for the relative neighbouring height, assuming that a taller tree creates more shading. Thus, the adjusted crown overlap ratio is given as:

$$\overline{overlap\ ratio}_i = \sum_{j=1}^{k} overlap\ ratio_{ij} \cdot \frac{height_j}{height_j + height_i}, \tag{3}$$

where k is a number of neighbours, while $height_i$ and $height_j$ are the heights of the reference and neighbouring trees, respectively. The crown light exposure is then calculated for each tree i at time t as:

$$cle_i = max[0, 1 - light\ loss\ multiplier \cdot \overline{overlap\ ratio}_i], \tag{4}$$

where *light loss multiplier* is set as a constant to 0.75.

2.3.4 Dieback

The dieback ratio indicates the percentage of tree that is dying, and/or it indicates the health condition of trees. For instance[1]: $dieback < 0.01 \Rightarrow$ *excellent condition*; $0.01 \leq dieback \leq 0.1 \Rightarrow$ *good condition*; $0.1 < dieback \leq 0.25 \Rightarrow$ *fair condition*; $0.25 < dieback \leq 0.5 \Rightarrow$ *poor condition*; $0.5 < dieback \leq 0.75 \Rightarrow$ *critical condition*; $0.75 < dieback \leq 0.99 \Rightarrow$ *dying condition*; $dieback = 1 \Rightarrow$ *dead tree*; The modelled dieback ratio depends on (i) the latest condition of the tree, (ii) the age via DBH and (iii) the health of neighbouring trees. The health of neighbouring trees determines the contagion risk, which is

[1] The class brackets are based on Nowak and Crane, 2002.

calculated as:

$$contagion\ risk_i = 0.9 * \sum_{j=1}^{k} \frac{\sum_{r=1}^{j} k^{r-1} \cdot dieback_r}{k^j}, \tag{5}$$

where k is a total number of neighbours j, and 0.9 is an adjustment parameter.

The new *dieback rate* is drawn from a uniform distribution between *healing range* and *dying range*, calculated as:

$$healing\ range_i = -1 \cdot (1 - contagion\ risk_i) \cdot \frac{\sqrt{dbh}}{M} \cdot healing\ rate_i, \tag{6}$$

where M is a parameter indicating maintenance scope, such that $M0 = 5$, $M1 = 4$ and $M2 = 1$; while $healing\ rate_i$ is a parameter set to 0.005. Furthermore, *dying range* is calculated as:

$$dying\ range_i = \frac{risk\ rate_i \cdot M}{(1 - contagion\ risk_i) \cdot \sqrt{dbh}}, \tag{7}$$

where *risk rate* is a parameter set to 0.005. Therefore, new dieback of tree i at time t is calculated as:

$$dieback_{it} = dieback_{i,t-1} + U \sim (healing\ range_{it}, dying\ range_{it}) \tag{8}$$

At every period, t there is a positive probability that a tree dies. It is modelled as a step function, and it is increasing in dieback ratio, i.e., the worse the condition, the higher the probability of dying; as well as the function is decreasing in maintenance scope, i.g., the higher the maintenance scope the lower the probability of dying.

2.3.5 Growth

Annual tree diameter growth is estimated for each tree as:

$$diameter\ growth_{it} = standard\ growth_i \cdot \frac{ffd_t}{f\bar{f}d} \cdot cle_{it} \cdot (1 - dieback_{it})[cm] \tag{9}$$

where $standard\ growth_i$ is a location-species specific parameter which is provided in the allometric database (see section 2.1.3) for each species and location, and it is constant during the simulation. If missing, the moderate annual growth rate of 0.8382 cm is applied (see the literature on diameter growth for details, such as: Lorimer et al., 1992; Nowak, 1994, Smith and Shifley, 1984; DeVries, 1987; Nowak, 2020). ffd_t is the number of frost free days in the year t, while $f\bar{f}d$ is the average number of frost free days at a given location.

The diameter growth rate is also adjusted based on the ratio between the current height of a tree and the average height at maturity for the species since, as the tree approaches to its maximum height, the growth rate decreases. The estimated tree height at maturity is derived from the literature and should be provided in the allometric database along with other parameters. The tree growth is then adjusted according to Nowak, 2020. When the height of the tree is more than 80% of its height at maturity, the diameter growth is linearly adjusted from 100% to 2.22% of growth when the tree reaches 125% of the maturity height. The growth adjustment of the tree i at time t is given as:

$$growth\ adjustment_{it} = 2.738 - 2.173 \cdot \frac{tree\ height_{it}}{maturity\ height_i} \tag{10}$$

For the trees that reach in between 80%-125% of their height at maturity, the diameter growth is given as:

$$\Delta dbh_{it} = diameter\ growth_{it} \cdot growth\ adjustment_{it}[cm] \tag{11}$$

New *dbh* of tree i at time t is given as:

$$dbh_{it} = dbh_{i,t-1} + \Delta dbh_{it}[cm] \tag{12}$$

The tree height is estimated using a location-species specific function provided in the allometric database, such as:

$$height_{it} = f_i(dbh_{it})[cm], \tag{13}$$

where the function f can be polynomial, exponential or, parametric with parameters α_i, β_i, and γ_i. If missing function, the height is estimated using the model of Fleming, 1988:

$$height_{it} = f(tree\ condition_{it}) \cdot 0.15[cm], \tag{14}$$

where the function f is a tree condition multiplier. If the tree condition is "excellent", "good" or "fair" the multiplier is 1; for "poor" it is 0.76; for critical 0.42; for dying 0.15; and it is zero for dead trees.

Tree *crown height* and *crown width* are also estimated using the location-species specific functions provided in the allometric database, selecting a particular functional forms as well as, parameters of the function that corresponds to the species and location. Both, *crown height* and *crown width* are functions of tree *dbh*.

2.3.6 Biomass

There are two types of equations in the allometric database currently available to estimate a tree's biomass. Type 1 is given as:

$$biomass_{it} = \frac{e^{\alpha_i + \beta_i \cdot \ln[dbh_{it}] + \gamma_i/2}}{1 - root\ to\ shoot\ ratio_i}[kg], \tag{15}$$

and type 2 is:

$$biomass_{it} = \frac{\alpha_i \cdot dbh_{it}^{\beta_i + \gamma_i}}{1 - root\ to\ shoot\ ratio_i}[kg], \tag{16}$$

where $root\ to\ shoot\ ratio_i$ is a factor that converts the predicted above ground biomass to whole tree biomass. It can be species-location specific and be provided in the allometric database, however, for the time being it is missing and set to 0.26, see Cairns et al., 1997. Parameters α_i, β_i, and γ_i are the parameters of the function and are provided in the allometric database as species-location specific parameters.

2.3.7 Sequestration

Carbon storage is estimated multiplying tree biomass by 0.5 (Chow and Rolfe, 1989 and Nowak, 2020). To avoid the overestimation of carbon storage for big trees, we set a carbon storage cap at $7500kg$, as in $iTree$, assuming carbon sequestration of $25kg/year$ for trees with carbon storage capped at $7500kg$. Therefore, the carbon storage is given as:

$$carbon\ storage_{it} = min[biomass_{it} \cdot 0.5, carbon\ cap][kg\ year^{-1}], \tag{17}$$

where $carbon\ cap$ is set to $7500kg$. If $biomass_{it} \cdot 0.5 < carbon\ cap$ the yearly carbon sequestration for tree i is calculated as:

$$carbon\ sequestration_{it} = carbon\ storage_{it} - carbon\ storage_{i,t-1}, \tag{18}$$

and otherwise, $carbon\ sequestration_{it} = 25[kg\ year^{-1}]$

2.3.8 Decomposition

The decomposition is a process of carbon dioxide release from dead trees or, if a tree is still alive, the decomposition may occur on its dead brunches due to diebacks or crown loss. The amount of carbon to be released due to dead branches of alive trees is calculated as:

$$CO2release\ alive_{it} = crown\ to\ trunk\ ratio_i \cdot dieback_{it} \cdot carbon\ storage_{it}, \tag{19}$$

where $crown\ to\ trunk\ ratio_i$ is a parameter set to 0.05.

The decomposition rate of dead brunches depends mainly on humans' actions, such as maintenance activities. For the time being, the model distinguishes two scenarios: (i) if the dead brunches are mulched and left to be decomposed in the nature; or (ii) if the dead brunches are immediately decomposed, i.g., burnt. The first scenario happens with 90% probability, and in this case the decomposition rate is given as:

$$decomposition\ rate_{i,t-\tau} = \frac{1}{k \cdot e^{\frac{t-\tau}{k}}}, \tag{20}$$

where parameter k determines the speed of the decomposition. It is set to 2 for mulched dead brunches, indicating the fast decomposition where in the first 3 years around 60% of carbon is released, corresponding to empirical findings for mulched biomass above ground. The index τ indicates a year when the brunches died. The release profile is then given as:

$$CO2release\ profile_{i,t-\tau} = decomposition\ rate_{i,t-\tau} \cdot CO2release\ alive_{i\tau} \quad (21)$$

When the dead branches are burnt, the amount of *CO2release alive* is immediately released.

When a tree dies, the entire sequestrated carbon dioxide is released in the years ahead. The decomposition of the tree's root and the biomass above ground are considered separately, since the decomposition of the root is independent of humans' actions. Therefore, the $CO2$ release from the two parts is given as:

$$CO2release\ root_{it} = root\ to\ shoot\ ratio_i \cdot carbon\ storage_{it}, \quad (22)$$

$$CO2release\ above\ ground_{it} = (1 - root\ to\ shoot\ ratio_i) \cdot carbon\ storage_{it} \quad (23)$$

where $root\ to\ shoot\ ratio_i$ is a percentage of tree's biomass under the ground.

There is 50% probability that a dead tree will be removed from the site. Given that it is removed, there is 70% probability that it will be burnt, in which case the release of $CO2$ is immediate and given as:

$$CO2release\ profile_{i\tau} = 0.7 \cdot CO2release\ above\ ground_{i\tau} \quad (24)$$

The rest 30% is assumed to be converted into sustainable products.

If the tree is not removed from the site, there is 40% chances that it will stay untouched (standing) and 60% probability that it will be mulched. In the first case, the slow decomposition rate is applied (see equation 20) for both root and above ground biomass, with parameter $k = 5$; while in the second case, the root is still decomposed by the slow rate while the biomass above ground is decomposed by the fast rate setting parameter $k = 2$. Therefore, the $CO2$ release profile of standing tree is computed as:

$$CO2release\ profile_{i,t-\tau} = decomposition\ rate^{slow}_{i,t-\tau} \cdot carbon\ storage_{i\tau} \quad (25)$$

The $CO2$ release profile of mulched tree is calculated as:

$$\begin{aligned} CO2release\ profile_{i,t-\tau} = {} & decomposition\ rate^{slow}_{i,t-\tau} \cdot CO2release\ root_{i\tau} \\ & + decomposition\ rate^{fast}_{i,t-\tau} \cdot CO2release\ above\ ground_{i\tau} \end{aligned} \quad (26)$$

The dead tree can also be replaced with a new tree which is planted at the location where the tree is dead. The replacement will depend on the maintenance project in place. For instance, if the maintenance scope is $M0$, there will not be any replacement, if it is $M1$ the tree is replaced with 30% probability, and in the case of maintenance scope $M2$, any dead tree will be replaced.

2.4 Outputs

The model provides two types of outputs: (i) primary variables, which are the result of ecosystem dynamics and are calculated during the simulation; (ii) secondary variables, which are derived using the primary output to estimate ecosystem services.

2.4.1 Primary variables

The primary variables can be tree specific as well as aggregates on the ecosystem level. Tree specific variables trace out the state of each tree at any point in time, t such as: species, dbh, height, crown height, crown width, canopy overlap ratio, crown light exposure, tree condition, dieback, biomass, annual gross carbon sequestration, carbon storage, CO2 release root, CO2 release biomass above ground for standing tree, CO2 release biomass above ground for mulched tree, CO2 immediate release of biomass above ground, and tree position on the digital grid. Therefore, any variable x is displayed for each tree i at any time, t such as x_{it}.

On the other hand, the model provides the following aggregated variables on the ecosystem level: carbon storage, annual gross carbon sequestration, carbon released, number of alive trees, number of trees given the condition (excellent, good, fair, poor, critical, dying, dead, replaced), gross carbon sequestration standard deviation. Any aggregated variable X_t at time t is calculated as a sum of variable x_{it} over all trees, such as: $X_t = \sum_i x_{it}$.

Therefore, the model provides two data sets. The first panel data of all trees over time, and the second, time series data set of aggregated variables over time. Using these two data sets, the model calculates secondary variables that will be used in the impact analysis presented in section 3.

2.4.2 Secondary variables

The model estimates Leaf Area (LA), Bark Area (BA), Leaf Area Index (LAI), Bark Area Index (BAI), and Plant Area Index (PAI) for each tree i at any time t taking into account the leaf-off-leaf-on seasons for deciduous trees. Following Nowak, 1996, the total tree leaf area is calculated using the equation based on crown parameters, such as:

$$\ln LA_{it} = \beta_0 + \beta_1 H_{it} + \beta_2 D_{it} + \beta_3 S_{it} + \beta_4 C_{it}, \tag{27}$$

where βs are species-location specific estimated parameters provided in the allometric database. For the time being, the database contains only the parameters given by Nowak, 1996. H_{it} is crown height [m] for tree i at time t, D_{it} is average crown diameter [m] for tree i at time t; S is the average shading factor for the individual species i at time t, representing the light intensity intercepted by tree crowns (McPherson, 1984). It is a species-location specific parameter taken from Nowak, 1996 and provided in the allometric database. C is based on the outer surface area of the tree crown, calculated as $C_{it} = \pi D_{it} \cdot (H_{it} + D_{it})/2$ (Gacka-Grzesikiewicz, 1980).

Bark Area is calculated as:

$$BA_{it} = dbh_{it} \cdot \pi(tree\ height_{it} - crown\ height_{it}). \tag{28}$$

Leaf Area Index is given as:

$$LAI_{it} = \frac{LA}{Under\ canopy\ area_{it}}, \tag{29}$$

where $Under\ canopy\ area_{it} = (\frac{crown\ width_{it}}{2})^2\pi$.

Bark Area Index is calculated as:

$$BAI_{it} = \frac{BA}{Under\ canopy\ area_{it}}, \tag{30}$$

while Plant Area Index is given as: $PAI_{it} = LAI_{it} + BAI_{it}$.

So far, the index t has represented a time unit of one year. However, to estimate the ecosystem services on an hourly basis, given the hourly weather data, the model merges the existing population data set with the hourly weather data where the index t becomes an hourly time unit. Moreover, the model corrects PAI for deciduous trees, taking into account the leaf-off-leaf-on seasons. The correction of PAI is implemented following Wang et al., 2008b, such as:

$$PAI_{it} = \frac{PAI_{it}^{max} - PAI_{it}^{min}}{1 + e^{-37(day_a - day_b)}} + PAI_{it}^{min}, \tag{31}$$

where $PAI_{it}^{max} = PAI$ and $PAI_{it}^{min} = BAI$. For PAI in spring, day_a is the Julian day (day of year) of simulation and day_b is the half-way point between leaf-off and leaf-on. For leaf-off transition, day_a is the half-way point between leaf-on and leaf-off and day_b is the day of simulation. 0.37 is growth rate parameter.

3 Ecosystem services - Impact analysis

3.1 Biomass, carbon storage and sequestration

The estimation of trees' biomass, carbon storage and sequestration is presented in details in the previous section. These ecosystem services are calculated based on the projections of digital-twin tree growth. Nevertheless, the model also enables their calculation based on real data. One can import either cross-sectional or panel tree population data on a specific site and calculate trees' biomass, carbon storage and sequestration. The calculation is identical as it is described in the previous section. Three biomass is calculated using either equations 15 and 16, while carbon storage and sequestration is computed using equations 17 and 18.

3.2 Water retention

The water retention module uses local hourly weather data and estimates the hourly contribution of an individual tree i to the avoided run-off and storm water retention. i is a tree identifier such that $i \in \{1, 2, 3, ..., N\}$, where, N is the total number of trees in the population. For the sake of readability, we drop index i from the equations in this section. However, the equations will contain the time index t, which identifies hour-day-month-year. The module allows a user to make scenarios with different weather conditions and simulate the impact over n years. For instance, one can simulate the impact of trees in extreme rain events, different temperature scales, etc.

The module is developed following existing literature on hydrology, mainly Shuttleworth, 1993, Hirabayashi, 2013, and Hirabayashi, 2015. Our aim is to trace out the additionality of trees, calculating their net impact. Therefore, we first calculate the total-gross contribution of trees. Then we design a hypothetical scenario without trees to capture a part of the evaporation dynamics that would occur anyway. Subtracting the hypothetical scenario outcomes from trees' gross impact, we obtain their net impact - additionality.

The entire precipitation-interception event can be decoupled into two processes: (1) the process by vegetation interception and evaporation from vegetation; and (2) the process below the vegetation, including evaporation from the ground and run-off or infiltration. For the time being, we will distinguish two land types such as the pervious cover where the infiltration occurs, and the impervious cover where the run-off occurs.

3.2.1 Hourly precipitation-evaporation process for vegetated cover

3.2.1.1 Process by vegetation

As it is described in Hirabayashi, 2013, this process captures the dynamics of precipitation and evaporation from vegetation. It includes three stages, (1) when the precipitation starts and the vegetation still has the capacity to retain water; (2) vegetation drip starts when the vegetation has no capacity to retain water any longer; (3) when the precipitation stops and the evaporation from vegetation continues until the vegetation dries up or new precipitation starts. Therefore, assuming that precipitation starts at time t we calculate the vegetation storage Sv_t[m] at stage 1, equivalent of water height in meters (m) over a plain surface, such as:

$$Sv_t = Sv_{t-1} + Pv_t - Ev_{t-1}. \tag{32}$$

If $Sv_t < 0$, Sv_t is set to zero, whereas if $Sv_t > Sv_{max}$, Sv_t is set to Sv_{max} and the first stage ends. Sv_{max} is calculated as:

$$Sv_{max} = S_L \cdot PAI, \tag{33}$$

where S_L is a specific leaf storage of water (=0.0002m), see Hirabayashi, 2013, and PAI is Plant Area Index [Leaf Area Index (LAI) + Bark Area Index (BAI)].

Precipitation (Pv_t)[m] that is retained by vegetation at time t is calculated as:

$$Pv_t = P_t - Pd_t, \tag{34}$$

where Pd_t[m] is precipitation drip through the canopy at time t. It is a part of precipitation that will hit the ground, passing through the canopy, despite available vegetation capacity to retain water. It is calculated as:

$$Pd_t = P_t(1 - c), \tag{35}$$

where c is canopy cover friction that depend on PAI:

$$c = 1 - e^{-\kappa PAI}, \tag{36}$$

where κ is an extinction coefficient (=0.7 for trees and 0.3 for shrubs), see Wang et al., 2008a. Therefore, the amount of rain hitting the ground during the first stage is equal to Pd_t[m].

Figure 2 illustrates the first stage in the precipitation process by vegetation. The precipitation has filled half of the canopy capacity, while a part of the rain Pd is passing through the canopy. Evaporation from vegetation at time t, $Ev_t[m]$ is calculated as:

$$Ev_t = \left(\frac{Sv_t}{Sv_{max}}\right)^{\frac{2}{3}} \cdot PE_t, \tag{37}$$

where PE_t is potential evaporation $[m]$ at time t calculated using the modified Penman-Monteith equation (Shuttleworth, 1993):

$$PE = \frac{1}{\lambda\rho_w}\left(\frac{\Delta R_n + \frac{D\rho_a c_p}{r_a}}{\Delta + \gamma\big(1 + \frac{r_s}{r_a}\big)}\right), \tag{38}$$

where R_n is the net radiation (available energy) $[MJ\ m^{-2}]$. λ is latent heat of vaporization $[MJ\ kg^{-1}]$ given as:

$$\lambda = 2.501 - 0.002361T, \tag{39}$$

where T is temperature $[^\circ C]$. ρ_w denotes density of water $[kg\ m^{-3}]$, and it is estimated as:

$$\rho_w = 999.88 + 0.018T - 0.0051T^2. \tag{40}$$

Dividing R_n by $\lambda\rho_w$ one can obtain an equivalent debt of evaporated water in [m].

D is vapour pressure deficit (=$e_s - e$)[kPa], where e_s is saturated vapour pressure $[kPa]$ given as:

$$e_s = 0.6108exp\left(\frac{17.27T}{237.3 + T}\right), \tag{41}$$

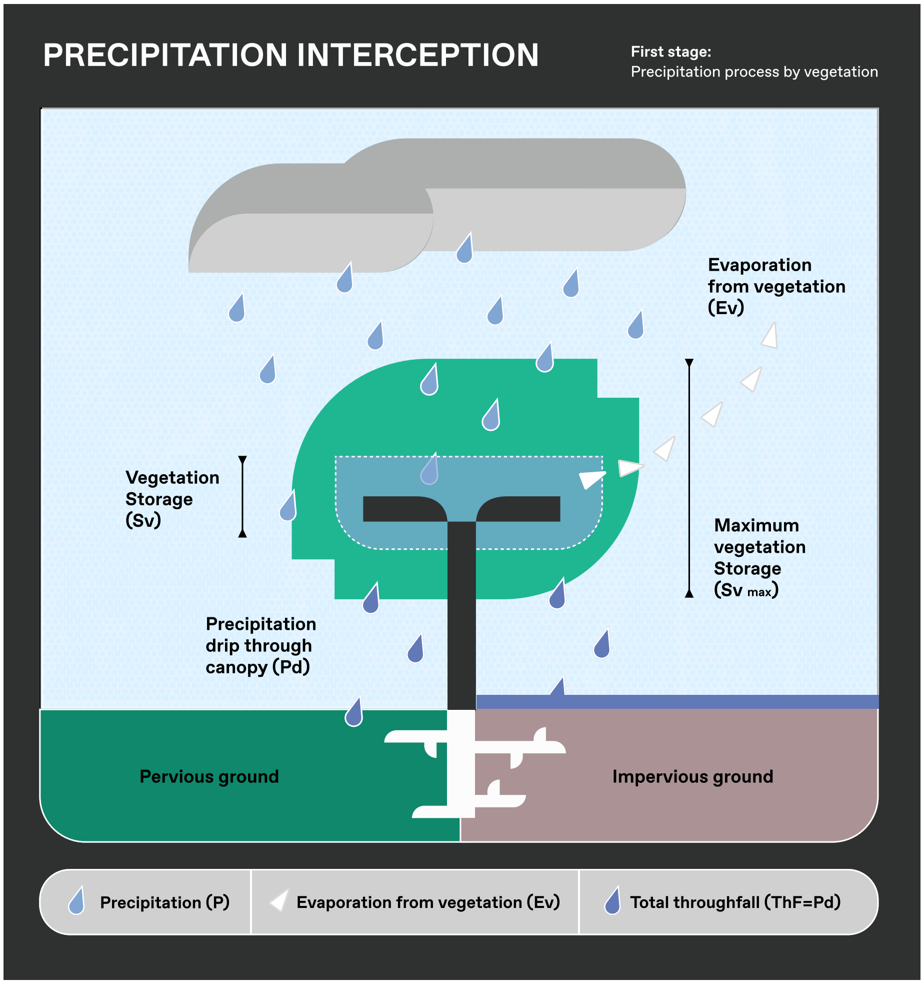

Figure 2: Trees AI, Precipitation interception phase 1.

and e is vapour pressure $[kPa]$ given as:

$$e = 0.6108exp\left(\frac{17.27DT}{237.3 + DT}\right), \tag{42}$$

where DT is dew point temperature $[°C]$. If $e_s - e < 0$, D is set to zero.

Δ is a slope of vapour pressure temperature curve $[kPa\,°C^{-1}]$ given as:

$$\Delta = \frac{4098e_s}{(237.3 + T)^2}. \tag{43}$$

ρ_a is the density of air $[kg\,m^{-3}]$ calculated as:

$$\rho_a = 3.486 \cdot \frac{P}{275 + T}, \tag{44}$$

where P is measured surface pressure [kPa].

c_p is specific heat of moist air (=1.013)$[kJ\ kg^{-1}\ {}^{\circ}C^{-1}]$.

r_a is aerodynamic resistance $[m\ s^{-1}]$ given as:

$$r_a = \frac{4.72 \cdot \ln\left(\frac{Z_t}{Z_{ov} d_t}\right)}{1 + 0.536 U_t}, \tag{45}$$

where Z_t is the wind estimate height for trees (tree height) [m]. Z_{ov} is mass transfer coefficient (=0.0123[m]). d_t is roughness height for trees (=0.95[m]). U_t is estimated wind speed at the tree top $[m\ s^{-1}]$ given as:

$$U_t = U \frac{\ln\left(\frac{Z_t}{d_w}\right)}{\ln\left(\frac{Z_u}{d_w}\right)}, \tag{46}$$

where U is measured wind speed $[m\ s^{-1}]$, Z_u is wind measurement height (=10[m]), and d_w is roughness height for water (=0.00137[m]).

γ is a psychometric constant $[kPa\ {}^{\circ}C^{-1}]$ given as:

$$\gamma = \frac{c_p P}{\epsilon \lambda} \cdot 10^{-3}, \tag{47}$$

where ϵ is the ratio of the molecular weight of water vapour to that for try air (=0.622).

r_s is stomatal resistance $[m\ s^{-1}]$ calculated as:

$$r_s = \frac{200}{PAI}. \tag{48}$$

Figure 3 demonstrates the second stage in the precipitation process by vegetation when the vegetation storage reach its maximum $Sv = Sv_{max}$. Precipitation that used to be retained by vegetation Pv now becomes vegetation drip Vd and together with the precipitation drip through the canopy Pd sums up to total throughfall ThF, thus $ThF = P$, while evaporation continues throughout this stage. Sv is calculated using equation 32 resulting in $Sv = Sv_{max}$, while Pv, Pd, and Ev are calculated using equations 34, 35, and 37 respectively.

For the first time in the second stage, where $Sv_{t-1} < Sv_{max}$ vegetation drip Vd[m] is calculated as:

$$Vd_t = Pv_t - (Sv_{max} - Sv_{t-1}) - Ev_t. \tag{49}$$

After that Vd is calculated such as:

$$Vd_t = Pv_t - Ev_t. \tag{50}$$

Therefore, the total amount of precipitation reaching the ground in the second stage at time t is given as $ThF_t = Vd_t + Pd_t$.

Figure 4 shows the third stage, when the precipitation stops and the evaporation from vegetation continues, which is calculated using equation 37.

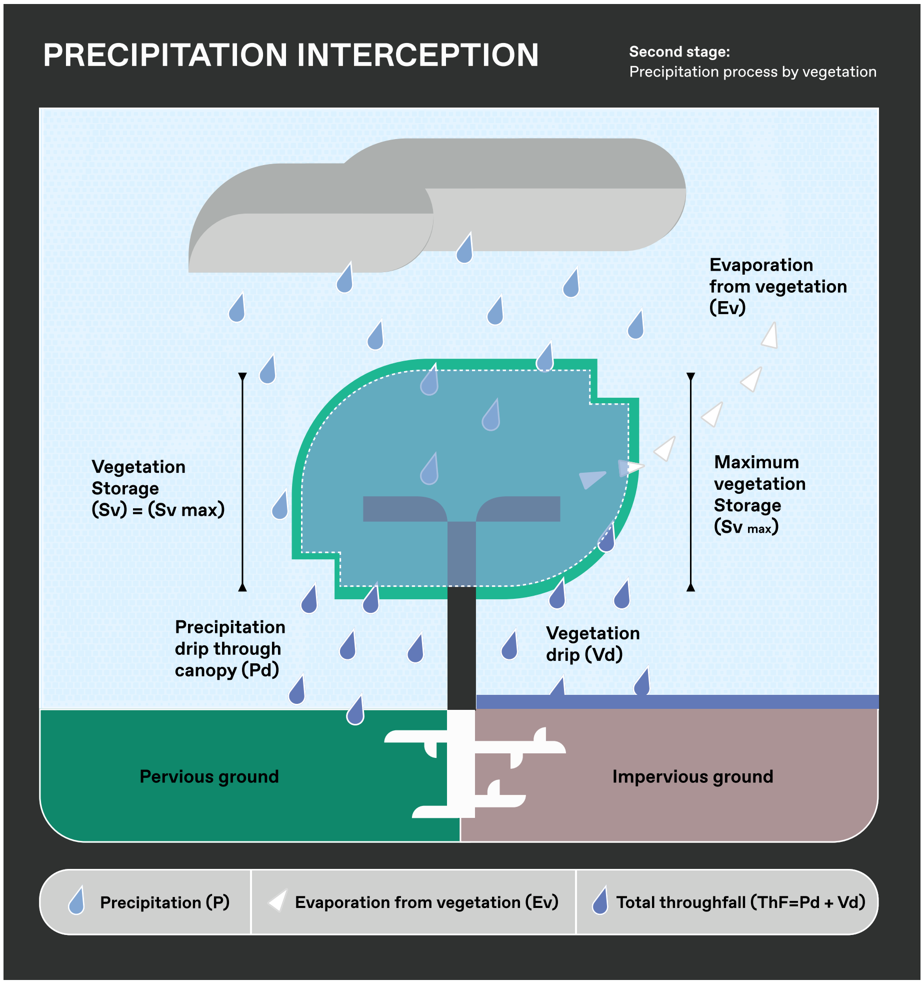

Figure 3: Trees AI, Precipitation interception phase 2.

3.2.1.2 Process by impervious cover under the canopy

Total precipitation reaching the ground ThF_t is a sum of precipitation drip through the canopy Pd and vegetation drip Vd after the vegetation storage reach its maximum capacity. Given this amount, we will apply here the same three stages as in the previous section. **Figure 5** illustrates the first stage of the process.

The impervious cover depression storage Svi_t [m] at time t is calculated as:

$$Svi_t = Svi_{t-1} + ThF_t - Evi_{t-1}. \tag{51}$$

If $Svi_t < 0$, Svi_t is set to zero, whereas if $Svi_t <= Si_{max}$, Svi_t is set to Si_{max} and the first stage ends. Si_{max} is constant (=0.0015m).

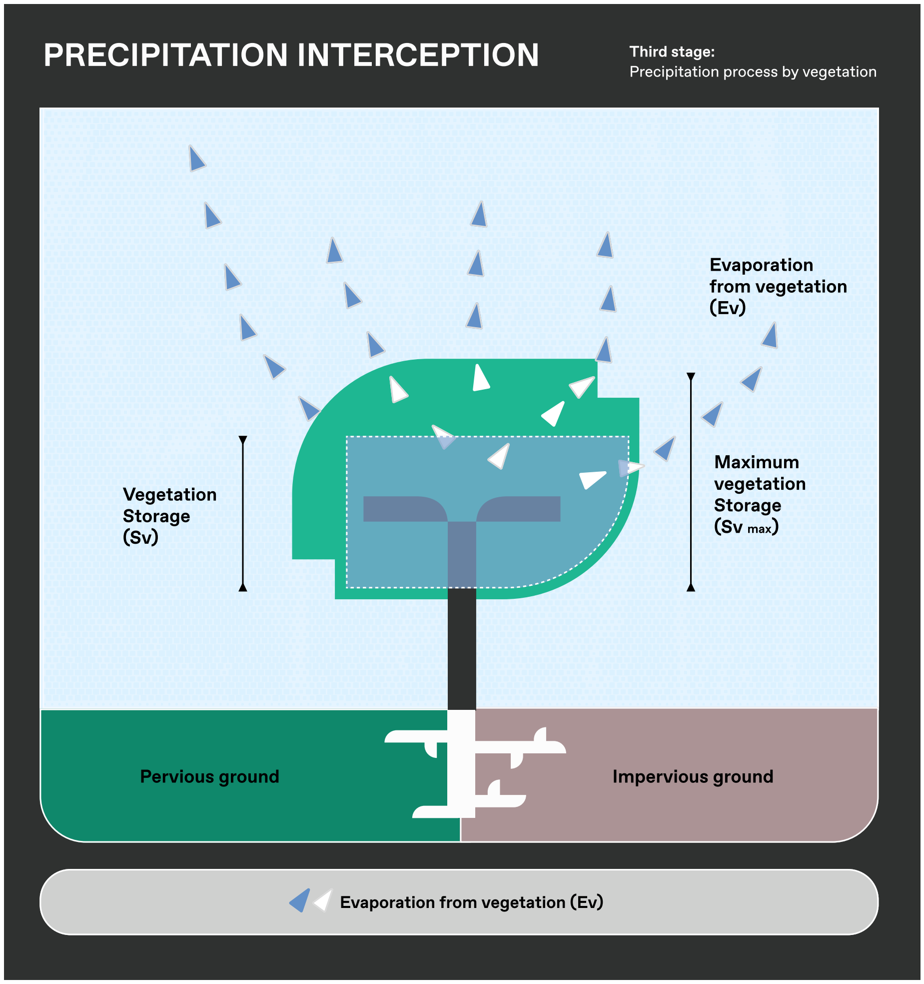

Figure 4: Trees AI, Precipitation interception phase 3.

Evaporation from the impervious cover Evi_t[m] at time t is calculated as:

$$Evi_t = \left(\frac{Svi_t}{Si_{max}}\right) \cdot PEg_t, \tag{52}$$

where PEg_t is the potential evaporation at time t from the impervious cover, calculated using equation 38.

Runoff over the impervious cover ROv_t[m] at time t is calculated as:

$$ROv_t = ThF_t - (Si_{max} - Svi_{t-1}) - Evi_t. \tag{53}$$

At the second stage, impervious cover depression storage Svi_t at time t is calculated using equation 51 having $Svi_t = Si_{max}$. Evaporation Evi_t is computed

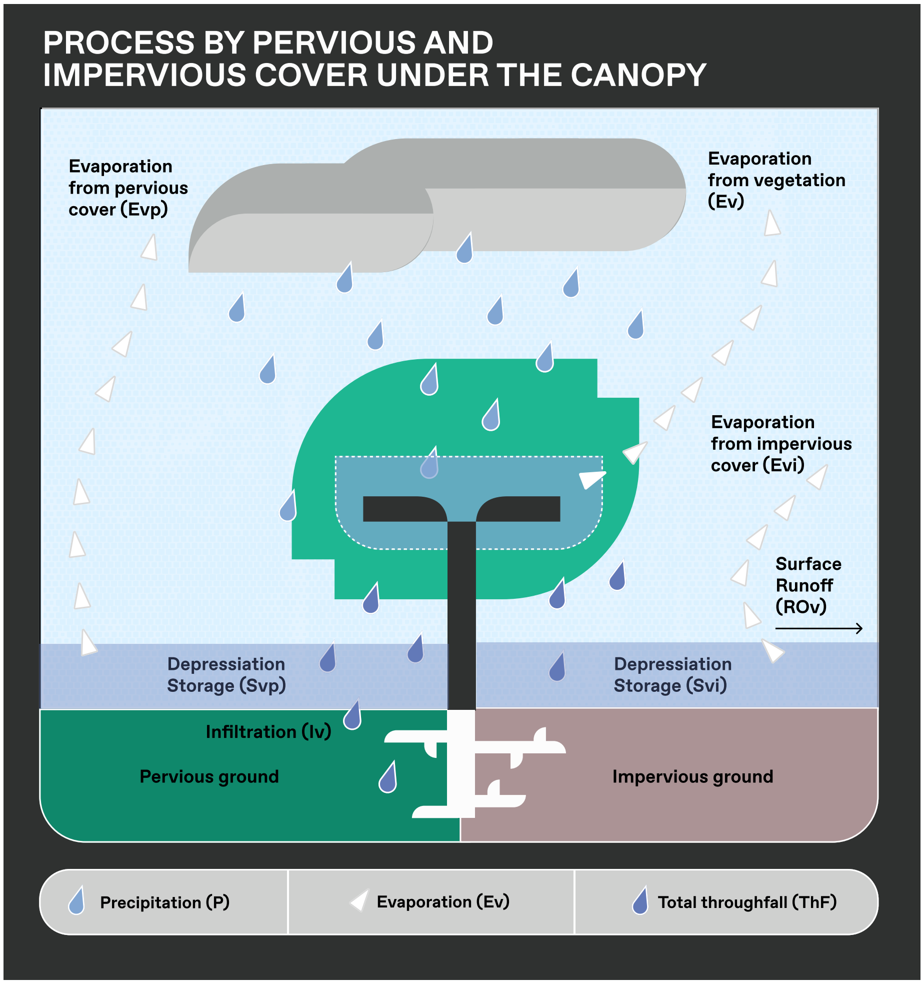

Figure 5: Trees AI, impervious/pervious covers under the canopy.

using equation 52, while runoff ROv_t at time t is calculated as:

$$ROv_t = ThF_t - Evi_t. \tag{54}$$

At the third stage, evaporation Evi_t is calculated using equation 52.

3.2.1.3 Process by pervious cover under the canopy

The same three stages as in the previous sections are applied to calculate the precipitation-evaporation process over the pervious cover under canopy. **Figure 5** illustrates the first stage of the process. The pervious cover depression storage Svp_t [m] at time t is calculated as:

$$Svp_t = Svp_{t-1} + ThF_t - Evp_{t-1}. \tag{55}$$

If $Svp_t < 0$, Svp_t is set to zero, otherwise if $Svp_t <= Sp_{max}$, Svp_t is set to Sp_{max} and the first stage ends. Sp_{max} is constant (=0.001m).

Evaporation from the pervious cover Evp_t[m] at time t is calculated as:

$$Evp_t = \left(\frac{Svp_t}{Sp_{max}}\right) \cdot PEg_t, \tag{56}$$

where PEg_t is the potential evaporation at time t from the pervious cover, calculated using equation 38.

Infiltration over the pervious cover Iv_t[m] at time t is calculated as:

$$Iv_t = ThF_t - (Sp_{max} - Svp_{t-1}) - Evp_t. \tag{57}$$

At the second stage, pervious cover depression storage Svp_t at time t is calculated using equation 55 having $Svp_t = Sp_{max}$. Evaporation Evp_t is computed using equation 56, while infiltration Iv_t at time t is calculated as:

$$Iv_t = ThF_t - Evp_t. \tag{58}$$

At the third stage, evaporation Evp_t is calculated using equation 56.

3.2.1.4 Transpiration by vegetation

Transpiration is the release of water in the form of water vapour from plants. The process is composed of two stages: (1) evaporation of water from cell walls, and (2) diffusion out of the leaf, mainly through stomata (Kramer, 1983 and Hirabayashi, 2015). Hourly transpiration flux $TF[g\,m^{-2}hr^{-1}]$ is calculated (following Kramer, 1983) as:

$$TF = \frac{C_{leaf} - C_{air}}{\frac{1}{g_s} + r_a} \cdot \frac{3600}{LAI}, \tag{59}$$

where g_s is stomatal conductance $[s\,m^{-1}]$ and $\frac{1}{g_s}$ is equivalent to stomatal resistance $r_s[m\,s^{-1}]$ given in equation 48. r_a is aerodynamic resistance given in equation 45. C_{leaf} and C_{air} are water vapour concentration at the evaporating surfaces within the leaf and water vapour concentration in the air $[g\,m^{-3}]$ respectively. They are calculated following (Monteith and Unsworth, 1990) as:

$$C_{leaf} = \frac{M_w e_s}{RT} = \frac{2165 e_s}{T}, \tag{60}$$

$$C_{air} = \frac{M_w e}{RT} = \frac{2165 e}{T}, \tag{61}$$

where M_w is the molecular weight of water $(= 18[g\,mol^{-1}] = 18000[g\,kmol^{-1}])$. R is universal gas constant $(= 8.314[J\,mol^{-1}\,K^{-1}] =$ $8.314[kPa\,m^{-3}\,kmol^{-1}\,K^{-1}])$. e_s and e are saturated vapour pressure and vapour pressure given in equations 41 and 42 respectively. T is temperature$[K]$.

The hourly transpiration flux mass per unit canopy cover, $TF[g\,m^{-2}\,hr^{-1}]$ is converted to depth $[m\,hr^{-1}]$ by multiplying 10^{-6} (1 g of water flux m^{-2} =

$10^{-6}\, ton\, m^{-2} = 10^{-6}\, m^3\, m^{-2} = 10^{-6}\, m$). It is then adjusted based on hourly potential evapotranspiration $PET[m\, hr^{-1}]$ from plants and inside the soil. We first calculate the average ratio between TF and PET using only observations when PET is larger than TF during the leaf-on season, such as:

$$\bar{R} = \frac{\sum_t (TF_t/PET_t)}{n}. \tag{62}$$

Then, when $TF > PET$ along the year or during the leaf-off season, TF is adjusted as:

$$TF = \bar{R} \cdot PET, \tag{63}$$

where PET is calculated as the sum of water released from plants (transpiration) and inside the soil (evaporation) that would take place if a sufficient water source were available. To estimate the hourly potential evapotranspiration, we employ equation 38 using the aerodynamic resistance $r_a = \frac{208}{U_t}$.

3.2.2 Hourly precipitation-evaporation process for the ground cover - hypothetical case

3.2.2.1 Process by impervious cover - hypothetical case

In the hypothetical case without any vegetation, the same process as in section 3.2.1.2 can be applied, except that now all precipitation P_t[m] at time t reaches the ground. **Figure 6** illustrates the precipitation-interception process by impervious cover.

The impervious cover depression storage Sgi_t [m] at time t is calculated as:

$$Sgi_t = Sgi_{t-1} + P_t - Egi_{t-1}. \tag{64}$$

If $Sgi_t < 0$, Sgi_t is set to zero, whereas if $Sgi_t <= Si_{max}$, Sgi_t is set to Si_{max} and the first stage ends. Si_{max} is constant (=0.0015m).

Evaporation from the impervious cover Egi_t[m] at time t is calculated as:

$$Egi_t = \left(\frac{Sgi_t}{Si_{max}}\right) \cdot PEg_t, \tag{65}$$

where PEg_t is the potential evaporation at time t from the impervious cover, calculated using equation 38.

Runoff over the impervious cover ROg_t[m] at time t is calculated as:

$$ROg_t = P_t - (Si_{max} - Sgi_{t-1}) - Egi_t. \tag{66}$$

At the second stage, impervious cover depression storage Sgi_t at time t is calculated using equation 64 having $Sgi_t = Si_{max}$. Evaporation Egi_t is computed using equation 65, while runoff ROg_t at time t is calculated as:

$$ROg_t = P_t - Egi_t, \tag{67}$$

At the third stage, evaporation Egi_t is calculated using equation 65.

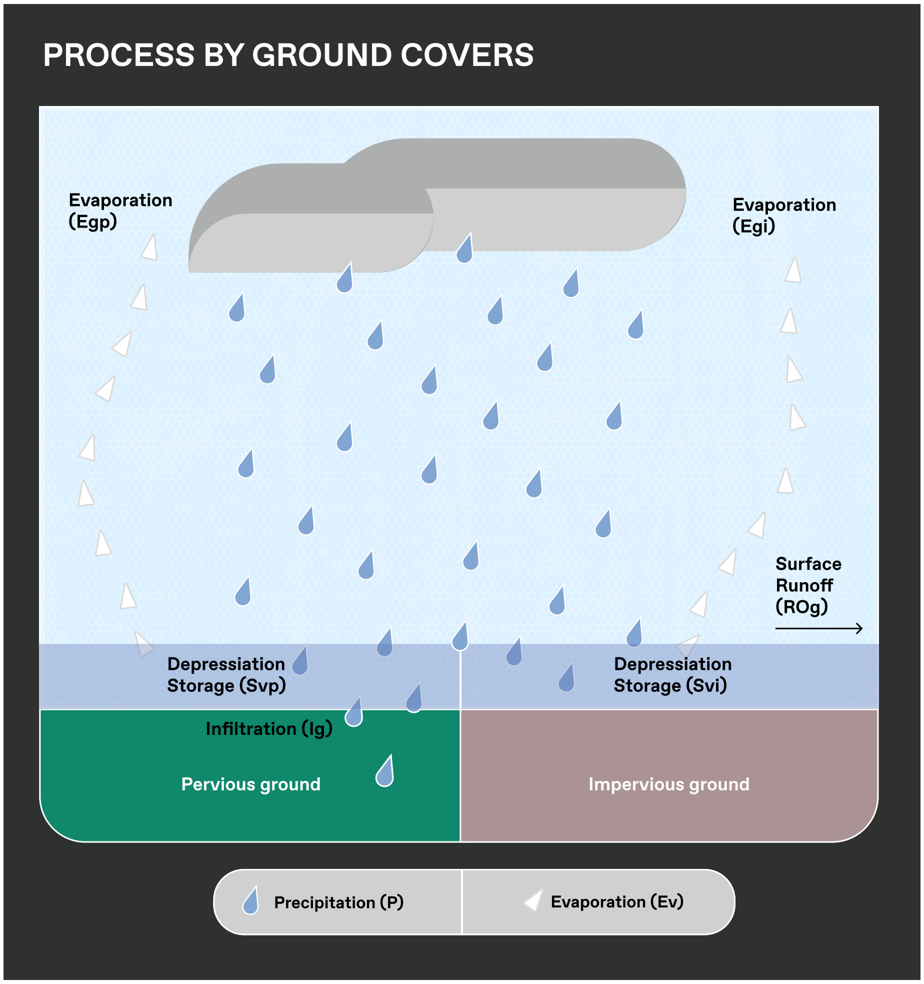

Figure 6: Trees AI, ground covers - hypothetical case.

3.2.2.2 Process by pervious cover - hypothetical case

Without vegetation, the same process as in section 3.2.1.3 can be applied, except that all precipitation P_t[m] at time t reaches the ground. **Figure 6** illustrates the precipitation-interception process by pervious cover. The pervious cover depression storage Sgp_t [m] at time t is calculated as:

$$Sgp_t = Sgp_{t-1} + P_t - Egp_{t-1}. \tag{68}$$

If $Sgp_t < 0$, Sgp_t is set to zero, whereas if $Sgp_t <= Sp_{max}$, Sgp_t is set to Sp_{max} and the first stage ends. Sp_{max} is constant (=0.001m).

Evaporation from the pervious cover Egp_t[m] at time t is calculated as:

$$Egp_t = \left(\frac{Sgp_t}{Sp_{max}}\right) \cdot PEg_t, \tag{69}$$

where PEg_t is the potential evaporation at time t from the impervious cover, calculated using equation 38.

Infiltration over the pervious cover Ig_t[m] at time t is calculated as:

$$Ig_t = P_t - (Sp_{max} - Sgp_{t-1}) - Egp_t. \tag{70}$$

At the second stage, pervious cover depression storage Sgp_t at time t is calculated using equation 68 having $Sgp_t = Sp_{max}$. Evaporation Egp_t is computed using equation 69, while infiltration Ig_t at time t is calculated as:

$$Ig_t = P_t - Egp_t, \tag{71}$$

At the third stage, evaporation Egp_t is calculated using equation 69.

3.2.3 Annual interception by trees

So far, we have calculated hourly interception for each individual tree and ground cover over n years, where index t identifies hour-day-month-year.

In the following, we will calculate annual outcomes, aggregating the variables on the yearly level. For instance, *annual evaporation from vegetation* is calculated as:

$$Ev_y = \sum_{month} \sum_{day} \sum_{hour} Ev_{h,d,m,y}, \tag{72}$$

where we replace index t with index y to identify annual outcomes. The same annual aggregation is applied to other variables as well.

3.2.4 Aggregate interception by trees

The interception by trees as well as by ground cover is defined in terms of the debth $[m]$ or in thousands of litters per square meter $[m] = [m^3/m^2] = [1000l/m^2]$. To calculate the aggregate benefits of a particular project, we will first calculate the total contribution of individual trees, multiplying the benefits $[m]$ by the under canopy area $[m^2]$, and then aggregate over all trees. Thus, the aggregate benefits will be converted to volumes $[m^3]$. Our main variables of interest will be: (1) *canopy interception loss*, which is identical to *evaporation from vegetation* Ev; (2) *storm water retention* SWR; (3) *improvement in storm water retention* SWR^{net}; and (4) *avoided runoff* RO^{net}.

Annual canopy interception loss per tree $Ev^{total}_{i,y}[m^3]$ is calculated as:

$$Ev^{total}_{i,y} = Ev_{i,y} \cdot UCA_{i,y}, \tag{73}$$

where $Ev_{i,y}$ is the annual evaporation from vegetation $[m]$ for the tree i in year y, and $UCA_{i,y}$ is the under canopy area $[m^2]$ of the tree i in year y.

Annual storm water retention per tree $SWR^{total}_{i,y}[m^3]$ is given as:

$$SWR^{total}_{i,y} = (Ev_{i,y} + TF_{i,y} + Evi_{i,y} \cdot ICS_{i,y} + Evp_{i,y} \cdot PCS_{i,y}) \cdot UCA_{i,y}, \quad (74)$$

where $Ev_{i,y}$ is the annual evaporation from vegetation $[m]$ for the tree i in year y, $TF_{i,y}$ is the annual transpiration $[m]$ for the tree i in year y, $Evi_{i,y}$ is the annual evaporation from the impervious cover under vegetation $[m]$ for the tree i in year y, $ICS_{i,y}$ is the impervious cover share $[\%]$ below the tree i in year y, $Evp_{i,y}$ is the annual evaporation from the pervious cover under vegetation $[m]$ for the tree i in year y, $PCS_{i,y}$ is the pervious cover share $[\%]$ below the tree i in year y, and $UCA_{i,y}$ is the under canopy area $[m^2]$ of the tree i in year y.

To calculate the annual improvement in storm water retention (additionality) of each tree, we first calculate the hypothetical annual storm water retention for the area below each tree $SWR^{total}_{hyp(i),y}[m^3]$, assuming that the trees were not there:

$$SWR^{total}_{hyp(i),y} = (Egi_y \cdot ICS_{i,y} + Egp_y \cdot PCS_{i,y}) \cdot UCA_{i,y}, \quad (75)$$

where $Egi_y[m]$ is the evaporation from the ground impervious cover in a year y, and $Egp_y[m]$ is the evaporation from the ground pervious cover in a year y. Thus, the annual improvement in storm water retention $SWR^{net}_{i,y}[m^3]$ per tree is calculated as:

$$SWR^{net}_{i,y} = SWR^{total}_{i,y} - SWR^{total}_{hyp(i),y}. \quad (76)$$

To calculate avoided runoff for each tree $RO^{net}_{i,y}[m^3]$ we calculate runoff per tree $RO^{total}_{i,y}[m^3]$ and runoff for the hypothetical case $RO^{total}_{hyp(i),y}[m^3]$ such as:

$$RO^{total}_{i,y} = ROv_{i,y} \cdot UCA_{i,y} \cdot ICS_{i,y}, \quad (77)$$

where $ROv_{i,y}[m]$ is runoff below the tree i in a year y. And,

$$RO^{total}_{hyp(i),y} = ROg_y \cdot UCA_{i,y} \cdot ICS_{i,y}, \quad (78)$$

where $ROg_y[m]$ is runoff for the hypothetical - ground cover case in a year y. Therefore, annual avoided runoff per tree is calculated as:

$$RO^{net}_{i,y} = RO^{total}_{hyp(i),y} - RO^{total}_{i,y}. \quad (79)$$

Once we have calculated the variables of interest for each tree, we can perform further aggregation per species, per genera, for the entire site, etc. Here we will present an example where we aggregate the outcomes for the entire project. Thus, the annual canopy interception loss for all trees is calculated as:

$$Ev^{total}_y = \sum_i Ev^{total}_{i,y}, \quad (80)$$

annual storm water retention for all trees is given as:

$$SWR^{total}_y = \sum_i SWR^{total}_{i,y}, \quad (81)$$

annual improvement in storm water retention for all trees is calculated as:

$$SWR_y^{net} = \sum_i SWR_{i,y}^{net}, \tag{82}$$

and annual avoided runoff for all tree is given as:

$$RO_y^{net} = \sum_i RO_{i,y}^{net}. \tag{83}$$

3.3 Air pollution removal

The air pollution removal module uses location specific weather data and air pollutant measurements along with urban forest information to estimates the improvement in air quality in a specific location. The dry deposition of air pollution is estimated for particular matter less than 10 (PM10) following mainly Hirabayashi et al., 2015.

The module consists of several functions that enable us to estimate the improvement in air quality. Each function is described in the following.

3.3.1 Air pollutant flux calculation for PM10

The pollutant flux $F[g\ m^{-2}\ h^{-1}]$ for PM10 is given as a product of deposition velocity $Vd[m\ s^{-1}]$ and the air pollutant concentration $C[g\ m^{-3}]$:

$$F = V_d \cdot C \cdot 3600 \tag{84}$$

$$F_{min} = V_{d,min} \cdot C \cdot 3600 \tag{85}$$

$$F_{max} = V_{d,max} \cdot C \cdot 3600, \tag{86}$$

Where $V_{d,min}$ and $V_{d,min}$ are minimum and maximum deposition velocity, while F_{min} and F_{max} are minimum and maximum pollutant flux.

3.3.2 Deposition velocity calculation for PM10

As in Hirabayashi et al., 2015, deposition velocity for PM10 is calculated following minimum, maximum and average values provided by Lovett, 1994:

$$V_d = V_{d,PM10,avg} \cdot \frac{BAI + LAI}{BAI + LAI_{PM10}}, \tag{87}$$

$$V_{d,max} = V_{d,PM10,max} \cdot \frac{BAI + LAI}{BAI + LAI_{PM10}}, \tag{88}$$

$$V_{d,min} = V_{d,PM10,min} \cdot \frac{BAI + LAI}{BAI + LAI_{PM10}}, \tag{89}$$

where $V_{d,PM10,avg}$ is the average deposition velocity for $PM10$ ($= 0.0064[m\ s^{-1}]$); $V_{d,PM10,max}$ is the maximum deposition velocity for $PM10$ ($= 0.01[m\ s^{-1}]$); while $V_{d,PM10,min}$ is the minimum deposition velocity for $PM10$ ($= 0.0025[m\ s^{-1}]$). LAI_{PM10} is Leaf Area Index for particle deposition ($= 6$) and BAI and LAI are Bark Area Index and Leaf Area Index respectively.

3.3.3 Air quality improvement calculation for $PM10$

The hourly air quality improvement per unit tree cover $I_{unit}(\%)$ is calculated as:

$$I_{unit} = \frac{F}{F + M_{total}} \cdot 100, \tag{90}$$

where F is pollutant flux $[g\ m^{-2}\ h^{-1}]$. M_{total} is total air pollutant mass per unit tree cover $[g\ m^{-2}\ h^{-1}]$, which is calculated as:

$$M_{total} = H \cdot C, \tag{91}$$

where H is urban mixing height $[m]$, and C is air pollutant concentration $[g\ m^{-3}\ h^{-1}]$.

Hourly air quality improvement, for total tree cover area, $I_{total}[\%]$ is given as:

$$I_{total} = \frac{F \cdot \frac{T_c}{100}}{F \cdot \frac{T_c}{100} + M_{total}} \cdot 100, \tag{92}$$

where T_c is total tree cover in [%] in the area.

3.3.4 Air pollutant ($PM10$) concentration change

The module calculates the change in air pollutant concentration as:

$$\Delta C = \frac{C}{1 - \frac{I_{total}}{100}} - C \tag{93}$$

where ΔC is the air pollutant concentration change $[\mu g\ m^{-3}]$ for $PM10$, and C is the air pollutant concentration $[\mu g\ m^{-3}]$ for $PM10$.

4 Conclusion

This document provides the technical specification for the novel Green Urban Scenarios Framework (GUS). For the time being, we have developed the core simulation engine and built a flexible framework that will be easily extended in the future. We have also estimated several ecosystem services provided by urban trees, such as biomass, carbon storage and sequestration, water retention and air pollution removal. In the forthcoming period, TreesAI aims to extend the framework introducing more ecosystem services as well as to expand the analysis to new locations onboarding new cities and enlarging auxiliary databases, such as the site configuration, the population configuration, the allometric and the experiment configuration database.

Acknowledgement

The work that has led the implementation of GUS Framework is partially funded by Google.Org Impact Challenge on Climate, 2021.

We would like to thank Arianna Smaron from Dark Matter Labs, for the illustrations on impact assessments and Chloe Treger for her feedback on the architecture diagram.

References

Cairns, M. A., Brown, S., Helmer, E. H., & Baumgardner, G. A. (1997). Root biomass allocation in the world's upland forests. *Oecologia*, *111*(1), 1–11.

Chow, P., & Rolfe, G. (1989). Carbon and hydrogen contents of short-rotation biomass of five hardwood species. *Wood and Fiber Science*, 30–36.

DeVries, R. (1987). *A preliminary investigation of the growth and longevity of trees in central park. new brunswick, nj, rutgers university* (Doctoral dissertation). MS thesis.

Fleming, L. (1988). *Growth estimation of street trees in central new jersey. new brunswick, nj: Rutgers university. 143 p* (Doctoral dissertation). MS thesis.

Gacka-Grzesikiewicz, E. (1980). Assimilation surface of urban green areas. *Ekologia polska.= Polish journal of ecology.*

Hirabayashi, S., Kroll, C., & Nowak, D. (2015). I-tree eco dry deposition model descriptions. 2015.

Hirabayashi, S. (2013). I-Tree Eco Precipitation Interception Model Descriptions, 1–21.

Hirabayashi, S. (2015). I-Tree Eco United States County-Based Hydrologic Estimates. *Washington, DC: US Department of Agriculture, Forest Service*, 1–10.

Kramer, P. J. (1983). *Water relation of plants* (tech. rep.). Academic Press,

Lorimer, C. G., Frelich, L. E., & Nordheim, E. V. (1992). Forest-tree growth rates and probability of gap origin: A reply to clark. *Ecology*, *73*(3), 1124–1128.

Lovett, G. M. (1994). Atmospheric deposition of nutrients and pollutants in north america: An ecological perspective. *Ecological Applications*, *4*(4), 629–650.

McPherson, E. G. (1984). Planting design for solar control. *Energy-conserving site design. Washington, DC: American Society of Landscape Architects*, 141–164.

Monteith, J., & Unsworth, M. (1990). Principles of environment physics. edward arnold, new york. *Principles of environment physics. 2nd ed. Edward Arnold, New York.*

Nowak, D. J. (1994). Atmospheric carbon dioxide reduction by chicago's urban forest. *Chicago's Urban Forest Ecosystem: Results of the Chicago Urban Forest Climate Project*, 83–94.

Nowak, D. J. (1996). Estimating leaf area and leaf biomass of open-grown deciduous urban trees. *Forest science*, *42*(4), 504–507.

Nowak, D. J. (2020). Understanding i-tree: Summary of programs and methods. *General Technical Report NRS-200. Madison, WI: US Department of Agriculture, Forest Service, Northern Research Station. 100 p.*, *200*, 1–100.

Nowak, D. J., & Crane, D. E. (2002). Carbon storage and sequestration by urban trees in the usa. *Environmental pollution*, *116*(3), 381–389.

Shuttleworth, W. (1993). Evaporation in: Maidment, dr handbook of hydrology.

Smith, W., & Shifley, S. (1984). Diameter growth, survival, and volume estimates for trees in indiana and illinois. res. pap. nc-257. *US Department of Agriculture, Forest Service, North Central Forest Experiment Station, St. Paul, MN*, 10.

Wang, J., Endreny, T. A., & Nowak, D. J. (2008a). Mechanistic Simulation of Tree Effects in an Urban Water Balance Model [1]. *JAWRA Journal of the American Water Resources Association*, *44*(1), 75–85. https://doi.org/10.1111/j.1752-1688.2007.00139.x

Wang, J., Endreny, T. A., & Nowak, D. J. (2008b). Mechanistic simulation of tree effects in an urban water balance model 1. *JAWRA Journal of the American Water Resources Association*, *44*(1), 75–85.