

Contents

1	Introduction	3
1.1	Biodiversity Conservation	3
1.2	Analysis of Factors in the Data Sheet	3
2	Variables and Assumptions	5
2.1	List of Variables	5
2.2	Assumptions and Justification	5
3	The Stochastic Model	6
3.1	Objectives and Summary of the Stochastic Model	6
3.2	The Benefit of a Project with Respect to Time	7
3.2.1	Taxonomic Uniqueness	7
3.2.2	Threatened Level	7
3.2.3	Feasibility of Success	8
3.2.4	Ecological Benefits	8
3.2.5	Overall Score	8
3.2.6	Demonstration of Changes over Time	8
3.3	Algorithms and Optimization	9
3.3.1	The Implementation of Stochastic Process	9
3.3.2	Filtering by Benefit Value	10
3.3.3	Calculating an Estimation of Total Cost	10
3.3.4	Using Standard Deviation to Eliminate Outliers	10
3.3.5	Implementation of the Algorithm	11
3.4	Results and Analysis	11
4	Sensitivity Analysis	12
4.1	Changing Benefit Benchmark	12
4.2	Changing Taxonomic Uniqueness Coefficient	13
4.3	Changing Ecological Benefit Coefficient	14
4.4	Changing Threatened Level Coefficient	15
4.5	Conclusion of the Sensitivity Analysis	16
5	Strength and Weakness	16
5.1	Strength	16
5.2	Weakness	17
6	Conclusion	17
7	One-page Memo	18
8	Appendix	19
8.1	Python Code of the Stochastic Process	19
8.2	Works Cited	23

1 Introduction

1.1 Biodiversity Conservation

The declining biodiversity of natural habitats is an urgent global crisis to be addressed. In Florida, 48 different species of plants are currently imperiled and necessitate the implementation of a reliable conservation plan to recover the population of these species. However, the fundraising efforts are hindered due to the complex impacts associated with the deciding the starting year of each project. These impacts vary in their cost, duration, feasibility of success, benefit, and the taxonomic uniqueness of their respective target plant.

The Stochastic Model (S. Model) collects data of threatened level and the taxonomic uniqueness of a species, the duration, yearly cost, and the feasibility of success of the reservation project. The S. Model allocates each project into different time slots. It then analyzes the factors above based on certain assumptions below and scores each project given their respective starting time.

The S. Model has four objectives: funding all 48 projects, minimizing the total amount of funds, minimizing the difference between the costs of conservation each year, and reaching a target benefit value. The model divides the total cost of conserving the plant by the feasibility of success and obtains an estimated cost that theoretically ensures the success of the project. Since the longer the conservation project is postponed, the more endangered the plant is, and the less likely it is to save it. Thus, the feasibility of success of the conservation project decays over time.

To optimize the schedule, the S. Model tries to minimize the difference between the costs of conservation each year along with minimizing total fund to be raised. This is realized through a ranking system that considers both factors above.

A detailed description of algorithms is provided in Section 3.

1.2 Analysis of Factors in the Data Sheet

The incorporation of the following factors is critical for establishing an accurate model in order to maximize the benefit of the organization. An analysis of each factor is provided below:

Time: The most important part of the S. model is to determine when to start a conservation project. The time postponed to initiate a project is a dynamic variable to be determined by a fundraising plan, which affects the overall benefits and feasibility of the project by leaving the species unprotected. S. Model establishes a funding schedule that considers the effect of time on other factors.

The primary objective of the S. Model is to establish a funding schedule that considers the effect of time on other factors. The optimized plan will yield a schedule of the starting plan of each project to address the time factor.

Taxonomic Uniqueness: Taxonomic uniqueness accounts for how unique a species is in terms of its taxonomy. For an endangered species, the higher its uniqueness, the higher the overall potential benefit from saving this species as it scales in biological importance.

In the HiMCM scenario, the uniqueness factor is obtained from the “Taxonomic Uniqueness” column of the data spreadsheet.

Threatened Level: The extent to which a species is threatened is measured by its declining population. This value increases if a species’ protection program is postponed. This would heavily affect the feasibility of success for the conservation project of a species.

In the HiMCM scenario, the Threatened Level is obtained from the “benefit” column of the data spreadsheet. This value is estimated to increase over time if the species is not conserved.

Feasibility of Success: The feasibility of success of a project indicates the probability of a success conservation of a species after all funding actions are completed, which can be affected by the current endangered state of a species. Therefore, although the amount of funds is constant, the decrease of the feasibility of success has a severe impact on the total adjusted cost.

In the HiMCM scenario, the feasibility of success is determined by the “feasibility of success” column in the spreadsheet, which by default decreases exponentially over time. The more endangered a species is, the less probable a project is going to succeed.

Ecological Benefits of the Conservation Project: The ecological benefit of a project given its starting time. This is rather a complex measure that includes a variety of factors: how feasible a project is, Threatened Level, the ecological value of the project, and the benefit of the project to the entire habitat.

In the HiMCM scenario, the benefit of a project is obtained from the “benefit” column of the data spreadsheet.

Annual Budget of the Conservation Project: The annual budget of the conservation plan is the sum of the costs of proceeding projects in a particular year. This is vital to the fundraising organizations since they have to consider the practicality of budget distribution. Depending on the situation of each funding year, the annual budget should be kept relatively consistent. This would minimize the probability of failing to raise sufficient funds for a particular year with a high budget.

The S. Model uses the standard deviation of yearly cost to evaluate the fluctuation of annual budget. In order to provide a realistic funding schedule for the entirety of the organization, there should be as little fluctuation from year to year as possible.

2 Variables and Assumptions

2.1 List of Variables

Here are the list of variables occurring in the calculation of the model.

l_i	The length of the protection program for the i^{th} plant
s_i	The starting year of the project for protecting the i^{th} plant
$u_{i,0}$	The initial taxonomic uniqueness of the i^{th} plant
$T_{i,0}$	The initial threatened level for the i^{th} plant
$f_{i,0}$	The initial feasibility of success for the i^{th} plant
$x_{i,0}$	The initial ecological benefit for the i^{th} plant
r_u	The annual percentage increment of the taxonomic uniqueness of the i^{th} plant
r_t	The annual percentage increment of the threatened level of the i^{th} plant
r_b	The coefficient determining the effect of uniqueness on overall benefit
f_i	The actual feasibility of success of the i^{th} project starting at year s_i
u_i	The actual taxonomic uniqueness of the i^{th} plant starting at year s_i
$T_{i,t}$	The actual threatened level of the i^{th} plant at year t
bt_i	The overall score of the conservation project of the i^{th} plant with respect to its starting year
sch	A 2-d array containing the cost for the i^{th} project in j^{th} year, which is stored in $sch_{i,j}$
sch^a	A 2-d array containing the cost for the i^{th} project in j^{th} year adjusted by the feasibility of success, which is stored in $sch_{i,j}^a$
bt_i	The overall benefit of the i^{th} project
exp_pct_b	The accepting percentage range of overall benefit with respect to $bt_{i,max}$.
C_{total}^a	The sum of adjusted total cost
yc_j	The yearly cost of j^{th} year into the funding plan
dev_c	The standard deviation of yearly costs from a plan
N	The duration of the entire plan
$Score$	The final score of the plan regarding total cost and deviation of yearly cost

2.2 Assumptions and Justification

For the convenience of the calculation, and due to the limitation of available data, the S. Model assumes the following statements:

Assumption 1: The benefit of the corresponding project directly reflects the threatened level of a given species.

Justification: For the convenience of the model and the lack of given statistics, it is assumed that the given benefit is determined by the status of the species and its uniqueness.

Assumption 2: The threatened level of a given species undergoes an exponential increase over time.

Justification: Since the population of an unprotected endangered plant species degenerates overtime, it is assumed that the total benefit of a certain project declines as the starting time of the corresponding conservation project postpones. For the convenience of this model, an exponential decay is used to estimate the change of population status.

Assumption 3: Once a conservation project starts, it cannot be paused.

Justification: Since a break during the course of a proceeding conservation project can affect the result of this project, it is assumed that the conservation project cannot be paused.

Assumption 4: Once a conservation project starts, its relative benefit and feasibility of success will not change.

Justification: As all conservation methods focus on conserving, it is assumed that the mentioned projects would also conserve its relative amount of benefit and feasibility of success rather than letting it disintegrate along the process.

Assumption 5: It is only possible to determine whether a project succeeds after it is finished.

Justification: Due to the inefficiency of data, it is not possible to determine the outcome of the project during its proceedings. This is to say that the feasibility of success cannot be separated into annual components to calculate the estimated budget.

Assumption 6: The feasibility of success of a given conservation project decreases as the threatened level increases.

Justification: As previously stated in Assumption 2, the threatened level of a species increases over time. The difficulty of conserving a species increases along with higher threatened level. Thus, the feasibility of success would decrease accordingly.

Assumption 7: The lower the feasibility of success of a given project is, the higher the total cost.

Justification: Feasibility of success can directly affect the estimated total cost. It needs to be considered to ensure the funding for the project promises high successful probability. In order to quantify the expected cost for a given project, it is necessary to assume a quantitative relationship between them.

3 The Stochastic Model

3.1 Objectives and Summary of the Stochastic Model

With the complexity of the issue, the S. Model focuses on the following primary objectives:

1. Fund all 48 projects
2. Minimize the total cost of all projects
3. Minimize the fluctuation of yearly cost
4. Reaching a target benefit value

The fundamental concept of the S. Model is to filter funding plans based on their estimated overall benefits and rank them according to their the estimated total cost. In the first stage of filtering, the overall benefit is divided into different subsets of factors: uniqueness, threaten

level, ecological benefits, and feasibility of success. After the overall benefit of a funding plan is calculated, users need to provide a percentile value of the expected benefit with respect to the maximum benefit of all possible plans. Funding plans whose benefit is lower than this percentile are eliminated. Then, the remaining plans undergo another process of evaluation regarding the total cost of the plan and the deviation of yearly cost. This is achieved by multiplying the standard deviation of the yearly cost and an adjusted total cost to fund all project with respect to feasibility of success. The money required to fund a project is calculated by dividing the total cost of the project by its feasibility of success. It is higher if a project starts early in the process of the funding plan.

To fulfill the objectives, the S. Model takes into account all 48 projects in its random generation process of possible plans. The solution to the other three objectives are achieved by the characteristics of the algorithm. The specifics of algorithms are discussed in section 3.2 and section 3.3.

3.2 The Benefit of a Project with Respect to Time

The overall benefit of a project is determined by four major factors: The taxonomic uniqueness of a species, the extent to which a species is endangered, the ecological benefit yielded if the project succeeds, and the feasibility of success of a project. These factors are subjected to change over time. To address this issue, an algorithm is provided to evaluate the extent to which the change in these factors can affect the benefit outcome of the plan.

3.2.1 Taxonomic Uniqueness

According to section 1.2, the value $u_{i,0}$ is obtained from the "Taxonomic Uniqueness" column of the data spreadsheet. To enable further adjustment to the model, an exponential decay is used to express the relationship of the taxonomic uniqueness of a species to the starting time of its conservation project:

$$u_i = u_{i,0} (1 + r_u)^{s_i-1}$$

In the equation, r_u denotes for a annual increment of taxonomic uniqueness over time.

In the HiMCM scenario, the values are scaled to $[0,0.05]$ to adjust for the model. The default setting of the algorithm uses $r_u = 0.01$. Users can adjust this value to match the extent of change in the real world scenarios.

3.2.2 Threatened Level

The threatened level of a species reflects the state of which a species is endangered. (Section 1.2). In the HiMCM scenario, the initial value $T_{i,0}$ is obtained from the "benefit" column to indicate the Threatened Level. The Threatened Level undergoes an exponential increase if no measure is applied to conserve the endangered species (Section 2.2 Assumption 2):

$$T_{i,t} = T_{i,0} (1 + r_t)^{t-1}$$

Where $T_{i,0}$ is directly obtained from the "benefit" column (Section 2.2 Assumption 1). r_t is a weighting coefficient that determine the increasing Threatened Level over time. It is obtained from a region $[0,1]$, in which 0 indicates no change and 1 indicates the most amount of change. It is theoretically possible to estimate r_t if a species' population data in previous years can be obtained. However, the weight r_t is a parameter that reflects the user's understanding of the imperiled species. In the HiMCM scenario, r_t is scaled to $[0,1.25]$ with a default setting of $r_t = 1$.

3.2.3 Feasibility of Success

According to the assumption (Section 2.2 Assumption 6), the feasibility of success declines as the threatened level of a species increases. The actual feasibility of success with respect to the starting year of the project is calculated below:

$$f_i = f_{i,0} \cdot \frac{T_{i,0}}{T_{i,s_i}} = f_{i,0} \cdot \frac{T_{i,0}(1+r_t)^0}{T_{i,0}(1+r_t)^{s_i}} = f_{i,0}(1+r_t)^{-s_i}$$

The ratio between the initial threatened level and the threatened level when the project starts reflects the extent to which the feasibility of success will decrease over time. This is because if a species is more endangered, it is more unlikely for the project to succeed.

3.2.4 Ecological Benefits

To calculate the actual ecological benefit, one needs a more abundant data set that includes factors mentioned in 1.2. Due to the lack of information, the S. Model collects values of $x_{i,0}$ directly from the "benefit" column to represent the ecological benefit of a project. Since the ecological benefit of a project does not change over time, it is simply denoted as $x_{i,0}$.

3.2.5 Overall Score

The overall score of a project takes into account all the factors above. The formula of this calculation is given below:

$$bt_i = x_{i,0} \cdot f_i \cdot (1 - r_b u_i)^{s_i - 1}$$

Factor $x_{i,0}$ indicates that the higher the benefit of the project, the more need to initiate the project. Factor f_i indicates that a project with higher feasibility of success should be prioritized. Factor $(1 - r_b u_i)^{s_i - 1}$ is more complicated, since the longer the project is postponed, the less benefit it will produce. Therefore, the benefit of the project undergoes an exponential decay over time.

In the calculation, r_u is a user-generated weighting coefficient that indicates the extent to which the benefit of a project decays over time. It is obtained from a region $[0,5]$, in which 0 indicates no decay and 0.05 indicates the greatest decay. In the real calculation, r_u is scaled to adjust to the change of model outcome. In the HiMCM case, it is scaled to $[0,0.05]$ with a default value 0.01.

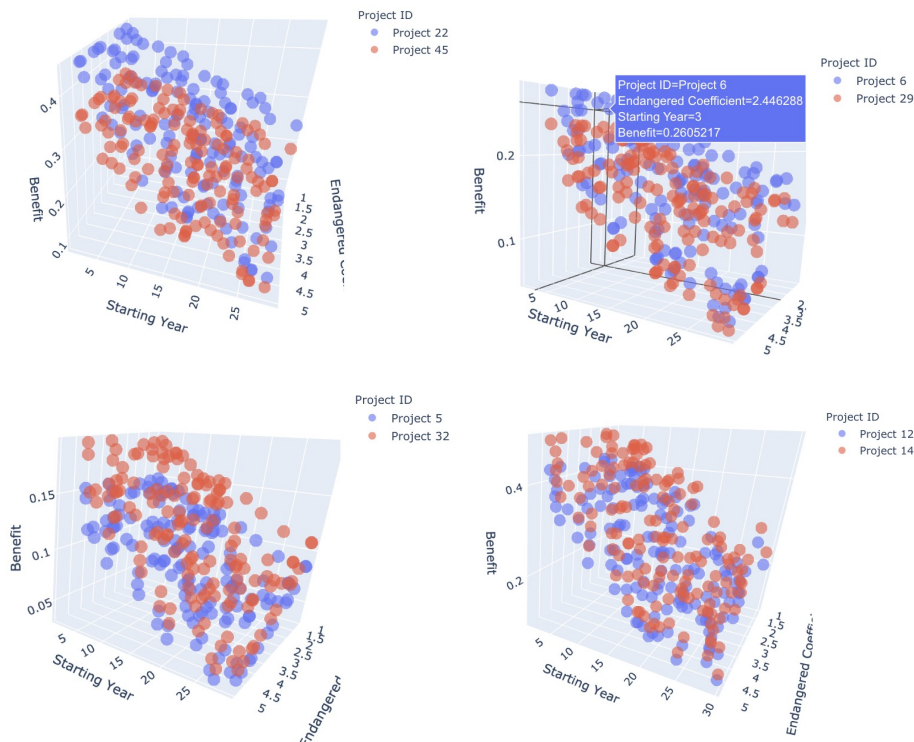
To obtain the overall benefit of a plan, it is simply to take the sum of all bt_i ;

$$bt = \sum_{i=1}^{48} bt_i$$

For the i th project, bt_i is maximized if the plan starts from year 1. Therefore, bt_i has a maximum if $s_i = 1$ for every $i \in [1, 48]$.

3.2.6 Demonstration of Changes over Time

This section provides some graphic information that demonstrates the relationship between some factors and time. The graph describes the relationship of the starting year of the project, the threatened level of the species, and the overall benefits in the end of the project. Eight projects are chosen for the analysis and are distributed in groups of two to compare the result of overall benefits.



These graphs show that as the starting year of a project postpones, the overall benefits of the project declines. Meanwhile, as the extent of population decay (Threatened Level) aggravates, the overall benefits also decreases.

3.3 Algorithms and Optimization

Since the goal of the model is to minimize fundraising while guaranteeing a certain level of outcome benefit, the S. Model incorporates these two factors to analyze whether a potential plan is feasible or not. Meanwhile, it is also important to consider that the fundraising schedule is reasonable to realize. This is to say that the fluctuation of annual budget needs to be minimized while fulfilling other requirements.

3.3.1 The Implementation of Stochastic Process

The S. Model is established upon the concept of Stochastic Process, a process that generates random funding plans and evaluates their total cost, deviation of annual costs, and benefit. For every project, a random number s_i is generated within a bound denoted by $uniform(lower, upper)$, where $lower$ indicates the minimum starting year and $upper$ indicates the maximum starting year. To accelerate the calculation of the process, given values of $upper$ and $lower$ are assigned to a set of projects within a given region:

l_i	s_i
$l_i \leq 9$	$uniform(1, 29)^*$
$9 < l_i \leq 15$	$uniform(1, 25)^*$
$l_i > 15$	$uniform(1, 9)^*$

*All upper and lower values in the chart are default settings. Real applications may be subjected to change.

After the random generation of s_i is completed, they are stored in the array “order” to reflect the chosen starting year of each project:

$$order[i] = s_i$$

The “order” array will then be used to calculate all outcomes of the plan described in the array.

3.3.2 Filtering by Benefit Value

Using the randomly generated starting year of the i^{th} project, bt_i , the overall benefit of the i^{th} project starting in year s_i is obtained. According to the previous definition of bt_i in Section 3.2.5, the overall benefit $\sum bt_i$ is maximized if every project starts in the first year. Although this scenario is not a realistic option, it is used as a benchmark to compare with the overall benefits of other plans. Then, the model requires a user to provide the minimum percentage of expected benefit for a funding plan, denoted as `exp_pct_b`, with respect to the benchmark maximum overall benefit. In this case, the benchmark of the model is $bm = (\sum bt_i)_{max} \cdot \text{exp_pct_b}$. If for the overall benefit of a funding plan $bt_i < bm$, then it is discarded, and the algorithm continues to generate the next set of s_i until $\sum bt_i \geq bm$ is obtained. The overall benefit of a feasible plan needs to be higher than a presupposed expectation of overall benefit.

After the eligible plans are filtered into a set P in which for every $order_x \in P$, $\sum_{s_i \in order_x} bt_i \geq bm$, the elements of the set P proceed into the next process of total cost evaluation.

3.3.3 Calculating an Estimation of Total Cost

Let $order_x \in P$, then for every $order_x$ a funding schedule sch is created. While sch only denotes for the original cost of i^{th} project starting in j^{th} year, an estimation of required cost regarding the feasibility of success needs to be generated. According to the assumption, the estimated cost of each plan is the original cost divided by the feasibility of success (Section 2.2 Assumption 7). Therefore, another funding schedule with adjusted cost is denoted as sch^a , in which:

$$sch_{i,j}^a = \frac{sch_{i,j}}{f_i}$$

For every i^{th} project, the cost of j^{th} year is calculated as the original cost divided by the feasibility of success in j^{th} year.

Notice that f_i is used regardless of the year j . This is because the feasibility of success for one project remains the same after it is initiated (Section 2.2 Assumption 3, Assumption 4, Assumption 5).

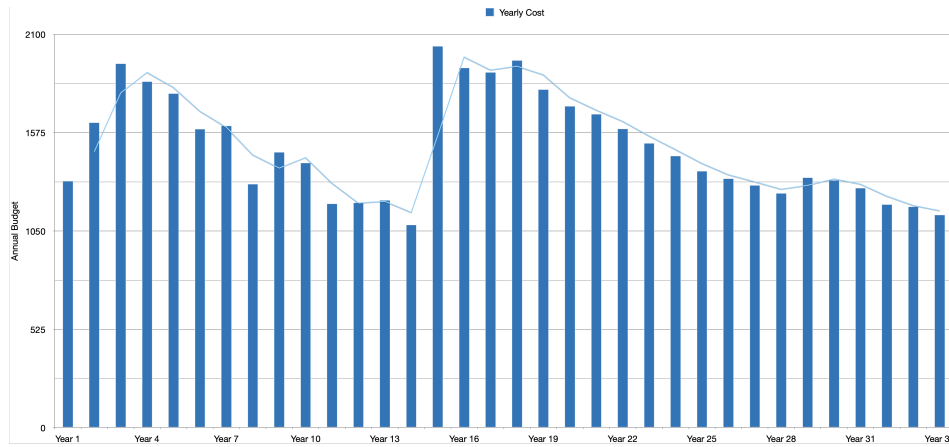
3.3.4 Using Standard Deviation to Eliminate Outliers

From sch^a two key factors are calculated: C_{total}^a the total cost of the plan, and dev_c the standard deviation of original yearly cost (the original yearly cost is the “solid money” that needs to be funded, whereas the estimated cost may consist of some “extra money” if the project succeeds).

The adjusted total cost of the project is obtained from:

$$C_{total}^a = \sum_{i,j} sch_{i,j}^a$$

For the deviation of original yearly costs, let yc be an array such that $yc_j = \sum_i sch_{i,j}$, which is the sum of the funding without adjustment for each year. Then, the standard deviation is taken from yc and yields dev_c :



The horizontal line indicates year, and the vertical line indicates yearly cost (in thousands of dollars)

The details of this plan regarding the starting time of each conservation project (marked in yellow), the original yearly costs for every project, and the yearly cost of the funding plan are provided in the chart above. The adjusted overall cost of this funding plan is \$107,134,718.35 including sufficient budget to theoretically ensure the success of all 48 projects. The yearly cost deviation of the best funding plan is \$267,411.23. The expected benefit is 89%. $C_{total}^a \cdot dev_c = 2.865 \times 10^{13}$

This optimal solution is filtered from a billion times of stochastic process using parameters $r_u = 0.01$, $r_b = 0.04$, $r_t = 1$ (scaled accordingly as mentioned in Section 3.2). The coefficient of benefit decays, and the susceptibility coefficient is high on their respective scales. The score for this funding schedule $Score_{optimal} = C_{total-optimal}^a \cdot dev_{c-optimal} = 28647631.6 = 2.9 \times 10^7$. The best scores obtained by randomly generated 10 million schedules is around 0.9×10^8 , as is shown in the sensitivity analysis. So our optimal funding schedule has a score less than a third of the best score values. An excel file of the best 3 schedules is obtained simply by the command: `get_schedule()`. This function by default does not require any input, and an excel file is created with key statistics of the schedule above.

4 Sensitivity Analysis

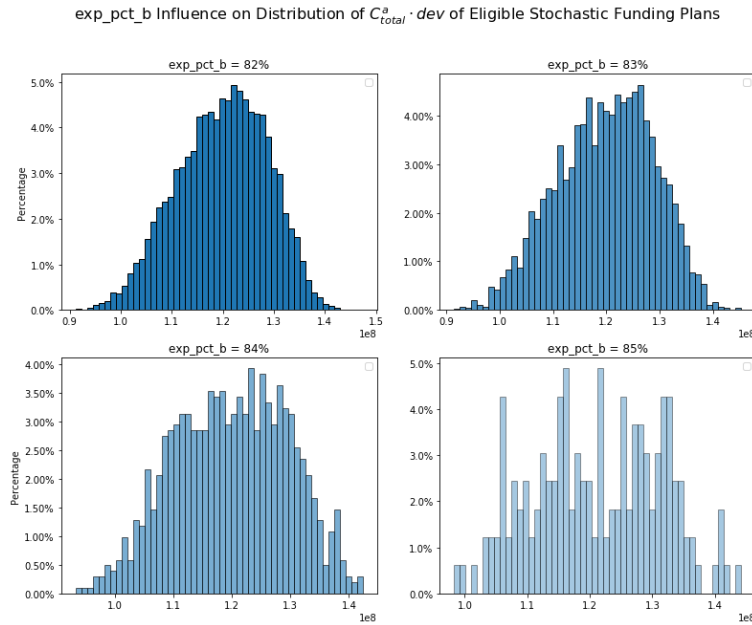
In this section, the S. Model is tested by some adjustments of its user inputs. The changes in results reflect how the changes in real scenarios and users' preference can modify the evaluation of possible plans. Four adjustments regards the benefit benchmark, the increase of taxonomic uniqueness over time for all species, the increase of Threatened Level for all species, and the effect of taxonomic uniqueness on the overall calculation of benefit.

For the unspecified factors, their values are taken as the default value identical to the values mentioned in section 3.3.5.

4.1 Changing Benefit Benchmark

This section analyzes how changes in the value of `exp_pct_b` alters the outcome of the algorithm. By adjusting the value of this factor, users can customize the required benefits from the estimated plan. Sometimes users may compromise a low benefit yield to decrease the overall cost of the plan; in other cases they may expect a higher benefit regardless the expense of the plan. The modification in `exp_pct_b` value reflects how users' preference will affect the result of the calculation. In theory, a higher value of `exp_pct_b` indicates a stricter requirement on the outcome of the funding plan. This would result in a larger percentage of possible plans eliminated in the calculation process.

For sensitivity testing, four different values of exp_pct_b are implemented in the model: 82%, 83%, 84%, and 85%. The algorithm in section 3.2 filters a set P containing plans that satisfy the exp_pct_b benchmark from all randomly generated possible plans. Then, the value $C_{total}^a \cdot dev_c$ for each feasible plan in P is obtained. The graph below investigates the distribution of $C_{total}^a \cdot dev_c$ for every plan that fits the requirement of an adjusted exp_pct_b value.



*Each process of random generation repeats for 10 million times.

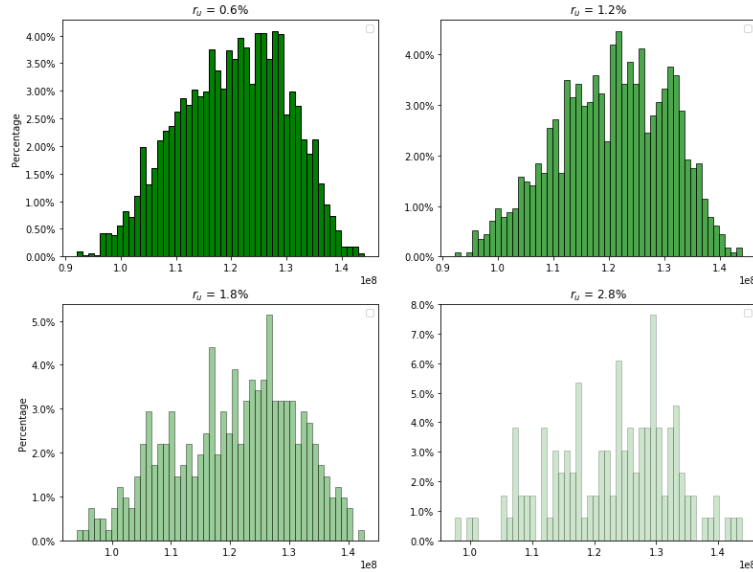
According to the graph, the distribution of $C_{total}^a \cdot dev_c$ value becomes more discrete as exp_pct_b increases. This is because the higher expectation of the overall benefit implies a smaller sample size that may vary in the overall score.

Meanwhile, the density of the region $[0.85, 0.93]$ decreases as exp_pct_b value increases. This shows that a higher restriction on the benefit of the project yields a higher overall cost and larger yearly cost distribution of a plan.

4.2 Changing Taxonomic Uniqueness Coefficient

This section analyzes how changes in the variable r_u alters the outcome of the algorithm. Changes in r_u determines how the user defines the change in taxonomic uniqueness over time. In theory, an increase in r_u indicates a more dramatic increase in the taxonomic uniqueness of a species if the starting time of its conservation project is postponed. This would result in a larger percentage of possible plans eliminated in the stochastic process. Since r_u may vary according to the circumstances of surrounding environment, users may need to adjust this value to better simulate the circumstances in which the model will be implemented.

For sensitivity testing, four different values of r_u are implemented in the model: 0.006, 0.012, 0.018, and 0.024. After the implementation of the adjusted r_u value, the algorithm in section 3.2 filters a set P containing plans that satisfy the exp_pct_b benchmark from all randomly generated possible plans. Then, the value $C_{total}^a \cdot dev_c$ for each feasible plan from P is obtained. The graph below investigates the distribution of $C_{total}^a \cdot dev_c$ for every plan in P that fits the requirement of a default exp_pct_b value.

r_u Influence on Distribution of $C_{total}^a \cdot dev_c$ of Eligible Stochastic Funding Plans

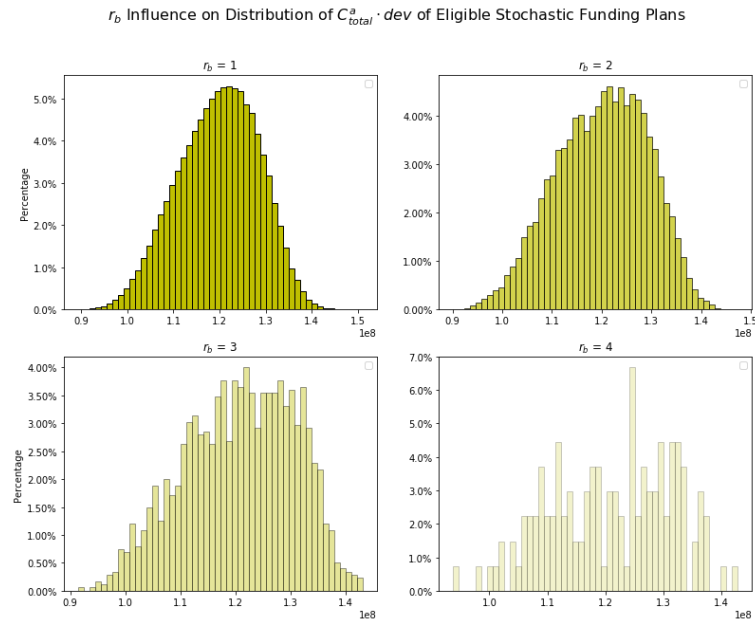
*Each process of random generation repeats for 10 million times.

According to the graph, the distribution of $C_{total}^a \cdot dev_c$ value becomes more discrete as r_u increases. This is because if the uniqueness of a species increases more dramatically over time, the overall benefit will decrease if the protection plan for this particular species is postponed. Therefore, a higher r_u value restricts the amount of time that a project can be postponed. This will lower the amount of feasible funding plans based on the algorithm's calculation, which is shown in the graph as a more discrete data distribution.

4.3 Changing Ecological Benefit Coefficient

This section analyzes how changes in the variable r_b alters the outcome of the algorithm. r_b determines the extent to which the overall benefit decreases over time before the start of the conservation project. In theory, an increase in r_b indicates a more dramatic decay in the benefit of the project if its starting time is postponed. This would result in a larger percentage of possible plans eliminated in the calculation process. In the real world scenario, this measure may be subjected to change due to the effect of multiple factors discussed in section 1.2. For the implementation of the S. Model, users may need to manually adjust r_b to suit the circumstances.

For sensitivity tests, four different values of r_b are implemented in the model: 1, 2, 3, and 4, in which a higher score suggests a more decayed benefit function. After the implementation of the adjusted r_b value, the algorithm in section 3.2 filters a set P containing plans that satisfy the `exp_pct_b` benchmark from all randomly generated possible plans. Then, the value $C_{total}^a \cdot dev_c$ for each feasible plan from P is obtained. The graph below investigates the distribution of $C_{total}^a \cdot dev_c$ for every plan in P that fits the requirement of a default `exp_pct_b` value.



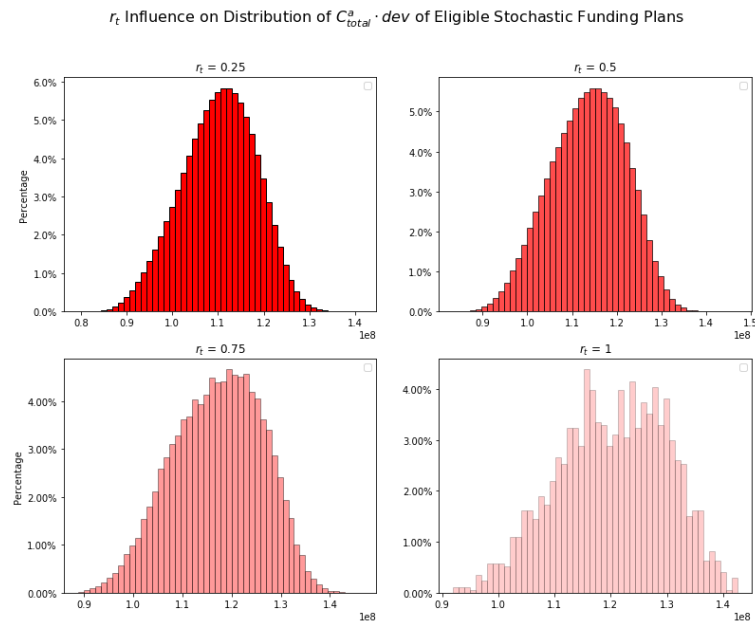
*Each process of random generation repeats for 10 million times.

According to the graph, the distribution of $C_{total}^a \cdot dev_c$ value becomes more discrete as r_b increases. This is because if the benefit of a species decreases more dramatically over time, the overall benefit will decrease if the protection plan for this particular species is postponed. In this case, a higher r_b value restricts the amount of time that a project can be postponed. This will lower the amount of feasible funding plans based on the algorithm's calculation, which is shown in the graph.

4.4 Changing Threatened Level Coefficient

This section analyzes how changes in the variable r_t results in the outcome of the algorithm. r_t determines how in real circumstances the population of an endangered species may decrease over time. In theory, an increase in r_u indicates a more dramatic increase in a species' Threatened Level if the starting time of its conservation project is postponed. This would result in a larger percentage of possible plans eliminated in the calculation process. Since r_t may vary according to the each species' distinct characteristics and conditions, users may need to manually adjust this value to better simulate the circumstances in which the model will be implemented.

For sensitivity tests, four different values of r_t are implemented in the model: 0.25, 0.50, 0.75, and 1.00, in which 1.00 suggests the most threatened level and 0 the least. After the implementation of the adjusted r_t value, the algorithm in section 3.2 filters a set P containing plans that satisfy the `exp_pct.b` benchmark from all randomly generated possible plans. Then, the value $C_{total}^a \cdot dev_c$ for each feasible plan from P is obtained. The graph below investigates the distribution of $C_{total}^a \cdot dev_c$ for every plan in P that fits the requirement of a default `exp_pct.b` value.



*Each process of random generation repeats for 10 million times.

According to the graph, the distribution of $C_{total}^a \cdot dev_c$ value becomes more discrete as r_t increases. This is because if the species becomes more endangered over time, the feasibility of success for the protection plan also decreases. This will lead to a lower overall benefit since the feasibility of success is directly proportional to the overall benefit. In this case, a higher r_t value also restricts the amount of time that a project can be postponed (for a larger postponing time yields a lower feasibility of success). This will lower the amount of feasible funding plans based on the algorithm's calculation, which is shown in the graph as a more discrete data distribution.

Meanwhile, the range of the data set tends to "translate" to the left as r_t increases. This is caused by the decrease on feasibility of success. Since a decreased feasibility of success requires a higher budget to ensure the success of a project, the overall cost of the project will therefore increase if the feasibility of success declines.

4.5 Conclusion of the Sensitivity Analysis

Based on the aforementioned four sensitivity tests, the S. Model proves to be helpful to yield results that are consistent with the theoretical estimations. This allows the S. Model to be flexible to simulate different real world conditions while maintaining the accuracy of its results.

5 Strength and Weakness

5.1 Strength

Based on the distinct features of the S. Model, it has several strength over other models.

1. Easy to Use

Compared to other models that required a plethora of data input and convoluted calculations, the operation of S. Model can be as simple as pressing a button. This is due to the fact that apart from the initial states of all species and their respective conservation projects, the operation of the S. Model does not require any other data from the user. All essential parameters of taxonomic uniqueness, Threatened Level, feasibility of success have a default value and can be modified if desirable. This largely shortens the learning curve while yielding the desirable results.

2. User Friendly Filtering System

The S. Model can be adjusted based on the user's preference. If the user prefers faster completion, the annual rate of increment of the endangered level and the decay rate of benefit can be increased. If the user focuses more on a steady annual budget, the score system can increase the importance of standard deviation.

3. Time-dependent Estimation

The S. Model takes into account how time affects the conditions of a species. By default the S. Model uses exponential decay function to estimate the change of benefit over time based on certain assumptions, which makes the model suited for simulating real world scenarios.

4. Pragmatic Funding Schedule

By minimizing the product of adjusted total cost and the standard deviation of annual budget, the model yields a realistic funding schedule with little fluctuation between years.

5.2 Weakness

Although the S. Model has many advantages, it still has some limitations in its real world application.

1. Risks of Failing Projects

The S. Model only gives the users an estimated benefit outcome. Although the probability of failure is inevitable, the S. Model cannot consider a scenario in which a project fails. That is, a re-funding of a failed project is not included in the algorithm.

2. Constant Parameter Values for Every Conservation Project/Species

Although r_u , r_t , r_b could be modified, they are the same for all projects once settled. This caveat could be fixed by allowing a list of r_u , r_t , r_b input for different projects, but this function is excluded for the sake of being user-friendly.

6 Conclusion

The Stochastic Model swiftly utilizes data provided by the spreadsheet to calculate and synthesize a plan that is as efficient as possible for the FRPCE board. The S. Model considers various aspects of imperiled plants, including their decreasing population and the increasing difficulty associated with protecting them. Then, it adjusts the actual benefit of the conservation projects by considering their respective time postponed, feasibility and target uniqueness to create a practical funding schedule that minimize the adjusted total cost as well as the annual standard deviation. It is user-friendly: three funding schedules will be created in an excel file with only one command "get_schedule()". Users can create their own functions on key variables and view interactive 3-D plots comparing different projects.

On the other hand, the S. Model does not calculate total funds required for the absolute success of every single project, therefore the exclusion of re-funding scenarios. However, the S. Model does provide a result of an adjusted overall cost to theoretically ensure the success of all projects.

Overall, the S. Model does a decent job achieving all the proposed objectives. The optimal funding schedule offers a practical reference for the FRPCE board to plan their projects.

7 One-page Memo

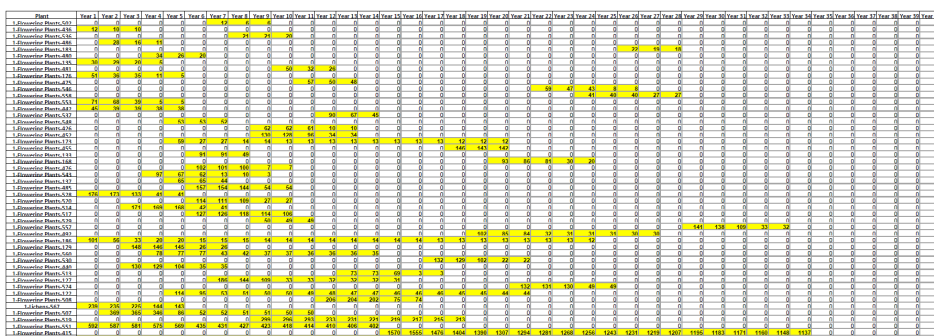
DATE: November 17, 2020 ‘

TO: FRPCE Board Members

FROM: HiMCM 2020 Team 10839

SUBJECT: A Recommendation on Biodiversity Funding Plan

Due to the need for a reliable and long-term funding schedule for the proposed 48 conservation plans, we propose that the Stochastic Model is constructed to optimize possible funding plans. Based on our calculation, a feasible funding plan is generated by its algorithm. The specific starting years and the yearly costs of each project according to this funding plan are given in the chart below:



All values of cost are in thousands of dollars

Based on the projection of this funding plan, all 48 recovery projects are allocated into 34 years. We expect the plan to yield 89% of its maximum benefit. The adjusted total cost of all conservation projects is \$107,135,000, and the standard distribution of yearly cost is \$267,411.

This plan provides a consistent annual funding budget, indicating that the FRPCE Board can expect higher probability of success with the presence of a reliable funding source.

The members of the FRPCE Board can make adjustments to several factors of the model regarding the uniqueness of each species, the rate of population decline for the imperiled species, feasibility of success of each project, and their ecological benefit. Also, in need for a higher percentage of benefit, the members of the FRPCE Board can raise the benefit benchmark to suit their preference.

Please be aware that the outcome of this model does not account for a certain level of risk in case of a failing project. The FRPCE Board needs to consider measures if a project fails in the end of its fundraising actions.

We are confident that the results of S. Model can provide a good reference to the further consideration of fundraising schedules. We sincerely hope that the future fundraising of all reservation projects will be successful.

Best,
Team 10839

8 Appendix

8.1 Python Code of the Stochastic Process

Below is the Python code for the calculation of S. Model.

```

1   import math
2   import random
3   import numpy as np
4   import pandas as pd
5   from numba import njit
6   import plotly.express as px
7
8   #data.xlsx must be in the same folder with this program
9   df = pd.read_excel('data.xlsx')
10  df.rename(columns={df.columns[0]: 'Plant',df.columns[2]:'Uniqueness'},
11            inplace=True)
12
13  L,TL = 48,24      #Reading data
14  x,u,f= np.zeros(L+1),np.zeros(L+1),np.zeros(L+1) #initial benefit,
15  uniqueness, feasibility
16  b_r,u_r = 4,0.01 #b_r: decay rate of benefit, u_r: increment of
17  uniqueness
18  cost = df.iloc[:48,4:4+TL].copy()
19  f[1:L+1],u[1:L+1],x[1:L+1] = df.iloc[:48,3],df.Uniqueness[0:48],df.
20  Benefit[0:48]
21  cost = cost.apply(pd.to_numeric, errors='coerce')
22  cost = round(cost/1000)
23  cost.replace(np.nan,0,inplace=True)
24  cv = cost.values #2d array of values in "cost" dataframe
25  leng=np.array([ 0,  3,  3,  3,  3,  3,  3,  4,  3,  5,  3,  5,  5,  5,
26  5,  3,  3,
27  5,  5, 16,  3,  3,  5,  5,  6,  3,  5,  5,  5,  5,  5,  3,  5,  9,
28  24,  5, 11,  5,  5,  5,  9,  5, 17,  5,  5, 10, 10, 14, 20])
29
30  @njit
31  def cur_u(initial_u,start_year,u_r): #current uniqueness
32  #u_r: annual increment of uniqueness without protection
33  return initial_u*(1+u_r)**(start_year-1)
34  @njit
35  def cur_t(initial_t,start_year,t_r): #current Threatened Level
36  #referred as the "endangered level"
37  #t_r is referred as the "endangered coefficient"
38  return initial_t*(1+t_r)**(start_year-1)
39  @njit
40  def cur_f(initial_t,start_year,t_r,initial_f,length):
41  #f_i = f_0 * t_s / t_e
42  t_s = cur_t(initial_t,1,t_r/100)
43  t_e = cur_t(initial_t,start_year,t_r/100)
44  return np.float32(initial_f * t_s/t_e)
45  @njit
46  def cur_b(initial_u,start_year,u_r, initial_b,b_r, initial_t,t_r,
47  initial_f,length):
48  #benefit of a project given its start_year
49  feasibility = cur_f(initial_t,start_year,t_r,initial_f,length)
50  return feasibility*initial_b*(1-cur_u(initial_u,start_year,u_r)*b_r/400)
51  **(start_year-1)
52
53  def details(order,u_r=0.01,b_r=4,t_r=1):
54  #return details given the starting years of each project, stored in
55  order

```

```

48     #order[i] stores the year when the ith project starts
49     bt_i,ben=0,0
50     for i in range(1,L+1):
51         bt_i = bt_i + np.float32(cur_b(u[i],1,u_r,x[i],b_r, x[i],t_r,f[i],leng[i]
    ]))
52     for i in range(1,L+1):
53         ben = ben + np.float32(cur_b(u[i],order[i],u_r,x[i],b_r,x[i],t_r,f[i],
    leng[i]))
54
55     sch = np.zeros((L+1,45))
56     sch_a = np.zeros((L+1,45))
57     for i in range(1,L+1):
58         for j in range(int(order[i]),int(order[i])+leng[i]):
59             sch[i][j] = cv[i-1,j-int(order[i])] #unadjusted cost
60             sch_a[i][j] = sch[i][j] / cur_f(x[i],order[i],t_r,f[i],leng[i])
61             #unadjusted cost
62             tc,yc = np.sum(sch_a),np.zeros(45)
63
64         for i in range(1,31):
65             yc[i] = np.sum(sch[:,i])
66             dev = np.std(yc[yc>0])
67         return {'tc':tc,'dev':dev,"yc":yc,"sch":sch,"sch_a":sch_a,'pct':int(ben/
    bt_i*100)}
68
69     @njit
70     def process(tot:int,u_r=0.01,b_r=4,t_r=1, exp\_pct\_b = 87):
71
72         bt_i = 0 #initial total benefit
73         col_tcdv=np.ones(tot)
74         col_tcdv = col_tcdv*300000000
75         for i in range(1,L+1):
76             bt_i = bt_i + np.float32(cur_b(u[i],1,u_r,x[i],b_r, x[i],t_r,f[i],leng[i]
    ]))
77
78         order,border,btcdv = np.zeros(49),np.zeros(49),3e8
79         bm_b = exp\_pct\_b/100*bt_i
80
81         for cnt in range(tot):
82             for i in range(1,L+1):
83                 if leng[i]<= 5:
84                     order[i]=int(math.ceil(random.uniform(0,24)))
85                 elif leng[i]<= 9:
86                     order[i]=int(math.ceil(random.uniform(0,19)))
87                 elif leng[i]<= 15:
88                     order[i]=int(math.ceil(random.uniform(0,14)))
89                 else:
90                     order[i]=int(math.ceil(random.uniform(0,6)))
91
92         ben = 0
93         for i in range(1,L+1):
94             ben = ben + np.float32(cur_b(u[i],order[i],u_r,x[i],b_r,x[i],t_r,f[i],
    leng[i]))
95
96         if ben<bm_b: continue
97
98         sch = np.zeros((L+1,45))
99         sch_a = np.zeros((L+1,45))
100        for i in range(1,L+1):
101            for j in range(int(order[i]),int(order[i])+leng[i]):
102                sch[i][j] = cv[i-1,j-int(order[i])]

```

```

103     sch_a[i][j] = sch[i][j] / cur_f(x[i],order[i],t_r,f[i],leng[i])
104
105     tc,yc = np.sum(sch_a), np.zeros(45)    #total cost adjusted, yearly cost
106     for i in range(1,31): yc[i] = np.sum(sch[:,i])
107     col_tcdv[cnt] = np.std(yc)*tc
108
109     if (tc*np.std(yc)<btcdv) & (tc!=0):
110         btcdv = np.std(yc)*tc
111         border=order.copy()
112         #print("          ", "trial complete")
113
114     return col_tcdv,np.min(col_tcdv),border
115     #collection of scores, minimum score, best funding schedule    order
116     that has the minimum score
117
118     def rep(tot:int, trials=100000, num=3, u_r=0.01,b_r=4,t_r=1, exp\_pct\_b
119           = 87):
120         col = np.ones((tot+1,49))*300000000    #collection of choices 1st filter
121         comp_c = np.zeros((num+2,49)) #second filter
122         col_tcdv = np.zeros(tot+1)    #store the "score = total adjusted cost *
123         std"
124         for i in range(1,tot+1):
125             nih,a,b=process(trials, u_r,b_r,t_r, exp\_pct\_b)
126             print(int(i/tot*100), '% complete')
127             col[i,:]=b
128             col_tcdv[i] = a
129             bm = np.percentile(col_tcdv[col_tcdv>0],int((num+1)/tot*100))
130             #benchmark bm: only return the top 3 funding schedules
131
132             cn=1
133             for i in range(1,tot+1):
134                 if col_tcdv[i]<bm: #minimal costs
135                     comp_c[cn]=col[i].copy() # copy "order" for optimal arrangements
136                     comp_c[cn,0],cn=col_tcdv[i],cn+1 # copy tcdv
137
138             return col_tcdv,comp_c
139
140     #compare two projects endangered coefficient-starting year-benefit
141     def plot_pj(idx1,idx2, L=150,u_r=0.01,b_r=4,t_r=1):
142
143         idx1,idx2=idx1+1,idx2+1
144         t_r,sim_o,ben1,ben2 = np.zeros(L),np.zeros(L),np.zeros(L),np.zeros(L)
145         cat1,cat2 = ['Project '+str(idx1-1),]*L, ['Project '+str(idx2-1),]*L
146         for i in range(L): t_r[i] = random.uniform(0,0.05)
147         for i in range(L): sim_o[i] = random.uniform(1,30)
148
149         for i in range(L):
150             ben2[i] = cur_b(u[idx2],sim_o[i],u_r,x[idx2],b_r, x[idx2],t_r[i],f[idx2
151             ],leng[idx2])
152             ben1[i] = cur_b(u[idx1],sim_o[i],u_r,x[idx1],b_r, x[idx1],t_r[i],f[idx1
153             ],leng[idx1])
154         df2 = pd.DataFrame({'Project ID':cat2,'Endangered Coefficient':t_r,
155             'Starting Year':sim_o,'Benefit':ben2})
156         df1 = pd.DataFrame({'Project ID':cat1,'Endangered Coefficient':t_r,
157             'Starting Year':sim_o,'Benefit':ben1})
158         Df = pd.concat([df1,df2],axis=0)
159         fig = px.scatter_3d(Df, x='Endangered Coefficient', y='Starting Year', z
160         ='Benefit',
161         size_max=5, color='Project ID',opacity=0.6)
162         fig.show()

```

```
157
158     def get_schedule(rep_tot=500, rep_trials=100000, num=3, u_r=0.01,b_r=4,
159     t_r=1, exp\_pct\_b = 87):
160     #automatically creates 3 optimal funding schedules in the current
    directory
161     best_tcdv,comp_c= rep(rep_tot, rep_trials, num, u_r,b_r,t_r, exp\_pct\_b
    )
162     writer = pd.ExcelWriter('Optimal Funding Schedules.xlsx', engine='
    xlsxwriter')
163
164     for cnt in range(1,4):
165     dfi,sch = comp_c[cnt].copy(),np.zeros((L+1,45))
166     dic = details(dfi)
167     sch = dic['sch']
168     dfi = pd.DataFrame(sch,index = df.index[:49].copy())
169     dfi.replace(np.nan,0,inplace=True)
170     dfi = dfi.astype(int)
171     dfi.drop(0,axis=1,inplace=True)
172
173     dfi.iloc[0,0] = "Standard Deviation of Expected AnnualFunding: "
174     dfi.iloc[0,1] = dic['dev'].astype(int)
175     dfi.iloc[0,2] = "Total Cost Adjusted to Feasibility of Success: "
176     dfi.iloc[0,3] = dic['tc'].astype(int)
177     dfi.iloc[0,4] = 'Percentage of Initial Benefit Achieved (ben/bt_i)'
178     dfi.iloc[0,5] = str(dic['pct'])+"%"
179     colname = [' ','']*len(dfi.columns)
180     for i in range(len(dfi.columns)): colname[i] = "Year "+str(i+1)
181
182     dfi.columns=colname
183     dfi.index = df.Plant[:49].shift(1)
184     dfi.drop(columns=dfi.columns[-4:],inplace=True)
185     dfi.rename(index={dfi.index[0]:'Key Stats'},inplace=True)
186     dfi.to_excel(writer, 'Sheet'+str(cnt))
187
188     writer.save()
189     return comp_c.astype(int)
190
191     get_schedule()
192     #Simply Shift+Enger, You get three optimal funding schedules created in
    one excel file, without requiring any inputs. To get better funding
    schedules, increase the default rep_tot=500 to 5000 or more.
```

Listing 1: Python Code: Execute the code in Jupyter Notebook

8.2 Works Cited

References

- [1] Conserving Habitat before It Is Too Late. Nov.17,2020, *Conserving Habitat before It Is Too Late.*, <https://scholarship.law.umt.edu/cgi/viewcontent.cgi?article=1424context=plrlr>.
- [2] The Difficulties of Conservation and Development. Nov.17,2020, *The Difficulties of Conservation and Development.*<https://wwf.panda.org/?199416/The-difficulties-of-conservation-and-development>.
- [3] Falk, Donald A., and Kent E. Holsinger. Conservation of Rare and Endangered Plants: Principles and Prospects. Nov.17,2020, *Genetics and Conservation of Rare Plants*, <https://pdfs.semanticscholar.org/77c8/8ae6ce5eae7bf623d5b1625472e5dcb8ae77.pdf>.
- [4] Lai, Shu-Mei, Wei-Chung Liu, and Ferenc Jordán. Nov.17,2020, *On the Centrality and Uniqueness of Species from the Network Perspective.*, <https://doi.org/10.1098/rsbl.2011.1167>.
- [5] What Makes a Species Endangered? Nov.17,2020, *What Makes a Species Endangered?*, <https://www.britannica.com/story/what-makes-a-species-endangered>.
- [6] Zhang, Wenhui, Yuangang Zhu, and Guobin Liu. Nov.17,2020, *Population Ecological Characteristics and Analysis on Endangered Cause of Ten Endangered Plant Species.*, Europe PMC. <http://europepmc.org/article/cba/389023>.
- [7] Negron-Ortiz, V. Nov.17,2020, *A Conservation Endowment for imperiled plants in Florida* [Conference session]., Botany 2019 Conference: Sky Islands and Desert Seas, Tucson, Arizona. The Botanical Society of America.