Zachary Schutzman

Economics 492 (Indep. Study)

Prof. Sahan Dissanayake

13 May 2015

Conservation Site Selection Modelling

Governments, land trusts, or private foundations may be interested in purchasing parcels of land for conservation purposes. If these purchasers are restricted by a budget and can only purchase a certain number of sites, the question of how to most effectively choose these sites naturally arises. A variety of methods can be used to select which sites should be conserved, depending on the objective of the conservators. Perhaps the organization will want to purchase as many sites as possible without regard to any features of those sites. In this case, the best strategy is to iteratively purchase the cheapest sites until the budget constraint is met. If the goal is to create something like a state park, the best strategy involves searching for a network of sites that is totally connected, and perhaps close to rectangular or circular in shape. In this project, the assumed goal was similar to a wildlife refuge, with a focus on maximizing the total number of species present in the network while also maximizing the connectedness of the sites selected.

In the model, the number of species in the system is maximized according to a set-covering condition. This is done as a way of representing a maximization of biodiversity in the system. Alternative metrics include maximizing the total number of populations conserved, which involves iteratively selecting the sites with the most species represented until the budget constraint is reached, or using a valuation method on the various species to determine how much conserving each species is worth, and then maximize that total. Because the total populations method may not conserve rare species and it is very difficult to place valuation on different animals, the set-covering metric will be the best choice in this simple model.

Additionally, ‘connectedness’ is a vague term, which can also be measured in a variety of different ways. The method used in this model is a minimization of total perimeter, which is actually phrased in the model as a maximization of shared edges between sites. While it provides different results, this metric is no more or less valid than minimizing pairwise distances between sites or minimizing the area of the region bounding all of the selected sites.

To test the model, a fictional landscape was fabricated. The landscape consisted of 400 sites in a 20 by 20 grid. Each site had at populations of at least one of 125 species, which were pseudorandomly distributed throughout the grid with probabilities relating to a beta distribution. Therefore, some sites had many species while others had few, and some species were present in many sites while others were present in few. An Excel sheet with conditional color formatting was used to generate visualizations of chosen landscapes (Appendix A).

A model where two features are maximized simultaneously is not an easy one to solve by hand, but computers are naturally suited to the task. This project used Gurobi, an industrial solver, and a Python language interface to reach solutions. Because the features of the sites and the network, such as shared edges and number of species present, are all integer values, and the site selection variables can be represented as “1” if a site is selected and “0” if it is not, a technique called ‘linear integer programming’ can be used to solve the model, and Gurobi handles these types of problems very well.

The mathematical statement of the model is:







However, Gurobi does not support simultaneously solving two different models simultaneously, which is effectively what happens when a model has two or more statements that require maximization or minimization of an expression. This necessitates rewriting the two maximizations above as:



where alpha and beta are weighting factors. In a model like this, it then makes sense to eliminate one of those factors and phrase the model as:



and let alpha range from zero to one. This then allows the model to be solved with a variety of different potential preferences, ranging from a model totally focused on species conservation and ignorant of connectedness to one that seeks the most connected system possible without regard to the biodiversity of the chosen sites.

One reason for using a computer to solve this problem is the sheer size of the solution space. A lower bound on the number of solutions for a given budget B is 400 choose , which for a modest budget of 40 is nearly 1020 possible landscapes. Even the computer cannot search the entire space at once. However, Gurobi uses a technique called ‘branching and bounding’ to find optimal solutions.

The basic premise of branching and bounding is that the computer looks at a partial solution and then looks through the possible solutions branching from that partial solution by traversing a tree. It will stop searching down any path once it comes across an infeasible solution and then returns the optimal solution from that particular subtree. After iteratively returning these partially optimal solutions, the computer can then look through this much smaller collection and return the best option, which is the true maximum of the objective function.

Additionally, it is of some interest to examine how good this model is at conserving rare species. Therefore, a rarity index value is assigned to each site. This value is equal to the sum of the reciprocals of the landscape-wide number of occurrences of the species present in a site. For example, if a site had representatives from species A, which appears 10 times in the whole system, species B, which appears 100 times, and species C which appears twice, the rarity index value for that site would be equal to . Clearly purchasing every site would have a rarity index equal to the total number of species in the landscape. The rarity index is calculated for each solution of the model but is not a factor in determining which sites to choose.

The above model was implemented in Python and solved for 20 budgets ranging from 5 to 100 and 33 alpha values ranging from 0.4 to 1. The resulting number of sites purchased, number of species conserved, total connectedness, and rarity index values were saved into a spreadsheet in order to generate some visualizations of the data.

The first chart (Appendix B) contains a series of curves reminiscent of the production possibilities curves from an introductory microeconomics course. Each colored curve represents a smooth extrapolation of the possible combinations of connectedness and species conserved for each given budget. This confirms the obvious tradeoff between connectedness and species conservation and suggests that at any given budget, the opportunity cost for either factor is not constant. Another interesting observation is that there are two logarithmic curves, one connecting the lower-most points of the curves, which are the solutions entirely focused on connectedness, and the upper-most points, which are the solutions entirely focused on species conservation.

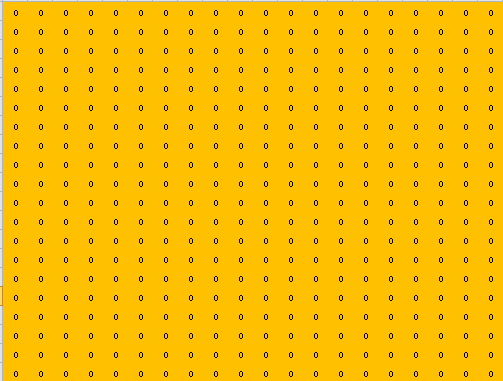
First, the shape of these curves implies that for any given value of alpha, there are decreasing marginal increases to the model’s objective function as the budget constraint increases, which is consistent with economic intuition and theory. The second observation is that both the horizontal and vertical distance between the curves is decreasing. In fact, these curves are indeed converging to a single point, which represents the possibility ‘curve’ generated by the single solution for the budget that allows purchasing the entire landscape. This convergence means that for any chosen value for either connectedness or species conserved, there is a lower and upper bound that can be placed on the values for the other. For a chosen value of connectedness, that bound is the vertical distance between the bounding curves, and for a chosen number of species conserved, the bound is the horizontal distance between the bounding curves. This matches the intuition that a model entirely focused on species will eventually have bordering sites if enough are chosen and a model focused entirely on connectedness will clearly conserve some populations present in the chosen sites.

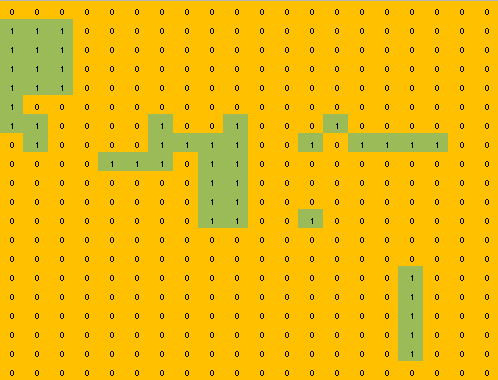
The next plot (Appendix C) is a graph of the objective value of the model against budget, with a color scale for the alpha values, with the cyan dots representing solutions weighted toward connectedness and magenta dots for solutions weighted towards species conservation. Again, if we connect all of the cyan dots with a curve and all of the magenta dots with a curve, we get two logarithmics which bound the solution space, albeit in a different way than the previous plot.

Notice that about two-thirds of the way through the plot, the magenta and cyan curves intersect. To the left of this point, which occurs around a budget of 67 in this landscape, the magenta dots have a higher objective value than the cyan dots, which reverses to the right of the intersection point. This implies that for lower budgets, a model more focused on species conservation has a higher value than one focused on connectedness, but the returns on that model diminish faster than those for a spatially focused model, to the point that for sufficiently high budgets, a model focused entirely on connectedness returns the highest objective value.

The final graph (Appendix D) is one that examines the rarity index against the budget. The color-coding is the same as the previous graph, with solutions entirely focused on species in magenta and connectedness in cyan. The first observation about this chart is that there is a strong general upward trend. This affirms that as more sites are purchased, the aggregate rarity of the species conserved increases. The more interesting quality about this plot is that, while the trend is positive, it is clearly not linear; rather, the general pattern is an S-shape, with the rate of increase higher at the far left and far right ends than in the middle. This may be because as the budget increases when it is initially very small, the number of species conserved rapidly increases, so the aggregate rarity would also likely increase rapidly. When the budget increases from an initially high value, it is possible that most or all of the species are already represented in a chosen site, so a site with few populations of rare species may be a good choice due to connectedness. The rate of increase for middle budgets may be comparatively lower due to the absence of these factors.

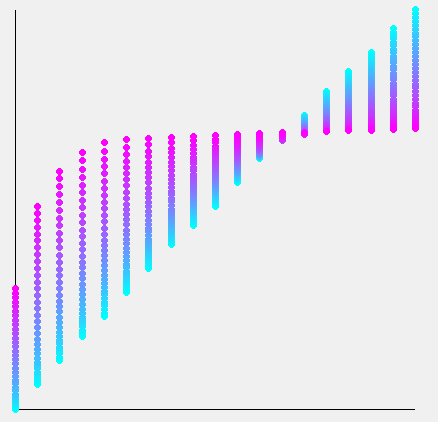
Overall, these preliminary results provide a great deal of insight to the techniques and results of conservation site selection models. It would be interesting to develop the other connectedness metrics, such as minimal pairwise distances or preset clusters to determine if the resulting graphs differ in any way. Because the landscape for this model was fabricated, it may not necessarily reflect the real-world applicability of the model. It would therefore be very useful to apply this model to an example of where a government or organization made these kinds of decisions to determine how strongly the results of this simple model align with real outcomes. Finally, valuation of a site comes from more than just the species present and the connectedness of the landscape. Factors like ecosystem services, natural features, and other potential uses of the land are important when deciding how conservation should occur. Potentially increasing the complexity of the model to incorporate some or all of these may or may not also provide different results.





Appendix A: Examples of a blank and a selected Excel visualization of the landscape

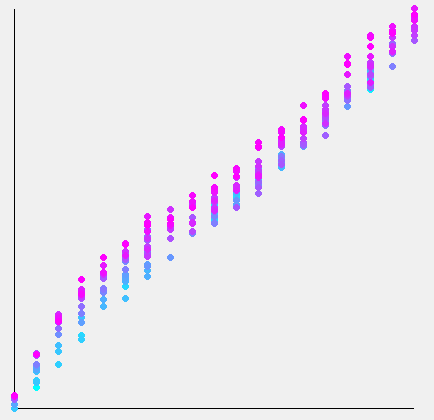
Appendix B: Possibilities curves for connectedness vs conservation



Budget

Objective Value

Appendix C: Plot of objective value vs Budget, colored by weight



Budget

Aggregate Rarity Index

Appendix D: Plot of Aggregate Rarity Index against Budget, colored by weight.