

Modeling of one-to-two year life expectancy of red harvester ant colonies based on neighborhood

Henry Cerbone and Matthew Smith

December 2022

Abstract

In this paper, we present initial, intuition building models for modeling the one-to-two year life expectancy of a population of red harvester ant colonies. The population, located near Rodeo, NM serves as a paradigmatic species for both collective behavior and group dynamics. Working from a dataset that is a result of a 35+ year census of population in question, we utilize both field collected data and environmental [14] in our study. Considering each colony as an individual organism, we construct various definitions of neighborhoods to help answer the biological question: what determines whether a new colony will survive from the first to second year? We utilize nearest neighbor, Voronoi, and other statistical tools in our approach. We also provide modeling to begin to elucidate the biological idea of a less favorable or “bad neighbor” through an *aggression metric*. We close with an exposition of both the mathematical and biological intuition behind the presented models.

1 Introduction

One of the best examples of collective behavior can be found in ants [3]. Particularly, we consider the red harvester ant *Pogonomyrmex barbatus* which lives in monogynous colonies. The harvester ants’ primary food source are seeds [10, 11] collected via daily foraging. A particular population of harvester ant colonies has been monitored in a longitudinal study over the last 35+ years in a research area outside of Rodeo, NM [13].

In this work, we consider each ant colony as an organism and ask the question: How does the foraging of neighbor colonies influence the one-to-two year life expectancy in new colonies? To help set the scene for this, we provide the survivorship curve in Figure 2. Note, there is a much less steep die off for young colonies than expected, however, considering that there is no metric of newly landed queens survival, this helps to explain the difference. This is a result of the mating style of harvester ants: new queens perform a nuptial flight to mate with males and then land at near random. This is not captured in the dataset used, but it has been shown that this has a steep die off [12].

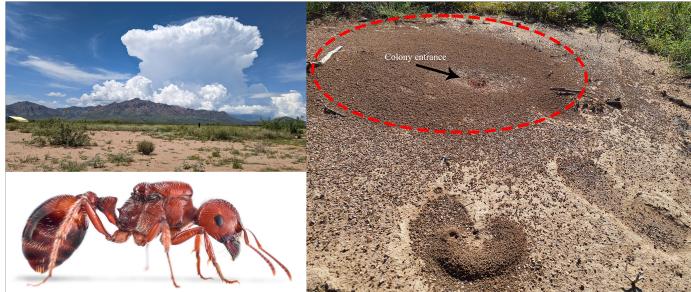


Figure 1: (Top left) the field site near Rodeo, NM; (bottom left) a lateral view of a harvester ant; (right) an example of a mid-year aged colony with associated colony entrance shown by the dashed red line.

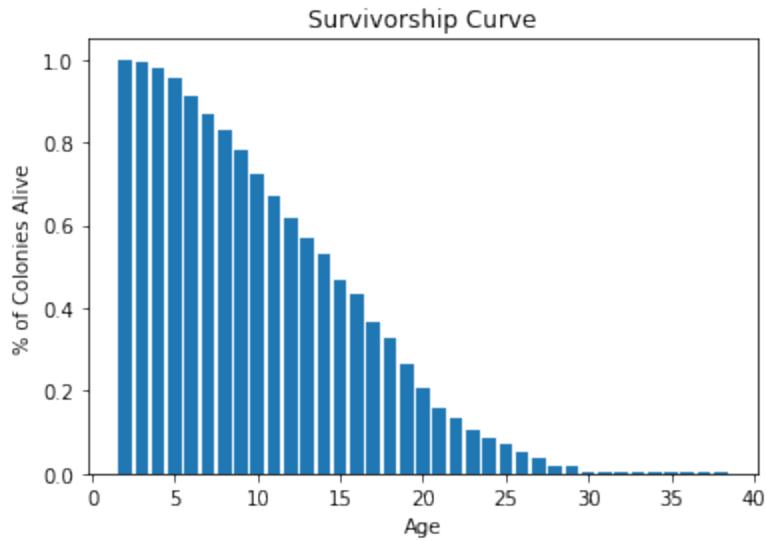


Figure 2: Cumulative survivorship of ant colonies by age

The foraging of a colony has been shown to be affected by many factors including site fitness [1], number of neighbors [13], and humidity [4]. Additionally, studies have been performed investigating the role of spatial distribution and size on fitness of colonies over time [7, 8].

In addition to the previously outlined results, past work has demonstrated the stratification of colonies by the effect that humidity has on their rate of foraging [4]. Factors such as these, in combination with heuristic field notes, show promise in modeling the interrelation of colonies. On this note, we also present initial mathematical underpinnings of an *aggression metric* which encapsulates the number of deaths of new colonies associated with a given neighbor. This

kind of thinking continues the precedent in the literature of considering intra-colony variations and their affect on both individual and neighboring colonies [8, 9, 6, 2].

2 Models

2.1 Data Preparation

First, we collected data from [13] on the ant colonies in the aforementioned longitudinal study. We store the longitude, latitude, age of colony, and whether they survive in the following year for each year y .

We also collected daily temperature and humidity data from San Simon, AZ from [14]. We chose San Simon, AZ because it has similar climate to Rodeo, NM and is less than an hour away. The same approach is used in [5]. We aggregated these daily measurements by taking the mean in each year and merged the resulting summary statistics with the colony data.

We split this data into a training and test set. The training data accounts for roughly 90% of the data and spans 1984-2016. The test data spans 2017-2020.

Finally, for 1-year-old colonies, we compute the average age of their neighbors using one of the methodologies discussed below.

2.2 Model Specifications

For each of the methods explored, we fit a logistic regression model on the training data to predict whether young colonies survive to the next year.

Since 1-year old colonies only die approximately 3% of the time, we fit each model using balanced loss. Furthermore, we evaluate each model on its balanced accuracy using leave-one-out cross-validation; while computationally intensive, this provides the best estimation of the robustness of a model on unseen data.

2.3 Baseline Models

We begin by considering a simple logistic model that models the probability of a young colony dying as a function of the age and distance of their nearest neighbor colony and the average humidity and temperature in that year. We denote these predictor variables as follows:

- D: Distance to nearest neighbor
- H: Average daily humidity
- T: Average daily temperature
- A: Age of the nearest neighbor

To determine the most appropriate basic logistic model for survival and establish a baseline by which we can assess future models, we calculate the leave-one-out CV balanced accuracy for every combination of these variables.

As shown in the Result section, the model considering only A and T yields the highest performance. Thus, we included these factors (or variations of them) in all additional logistic regression models we evaluated.

2.4 k-Nearest Neighbors - Definition

Building on our baseline models, we investigated whether the nearest neighbor truly has the largest influence on the survival of a young colony or if some other number of nearest neighbors k better reflects the viability of a neighborhood.

For a colony $i \in C$, we define its k -nearest neighbors N_i for some distance metric d (we use Euclidean distance) as follows:

Definition 2.1. $\forall j \in C$ s.t. $j \neq i$, $j \in N_i$ iff $d(i, j) \leq d(i, k)$ $\forall k \notin N_i$; $|N_i| = k$

2.5 k-Nearest Neighbors - Uniform Interaction

First, we assume that each of the k -nearest neighbors considered has the same level of interaction with the colony of interest. This assumption has some evidence in field studies but can be expanded in future work [8, 7].

Using the neighborhood definition above, we calculated the mean age of neighbors A using a standard average:

$$A_i = \frac{1}{|N_i|} \sum_{j \in N_i} a_j,$$

where a_j denotes the age of colony j .

Then, we performed grid search for the optimal number of neighbors k to consider by computing the leave-one-out CV balanced accuracy of the logistic regression model considering A and T.

2.6 k-Nearest Neighbors - Distance-Weighted Interaction

Next, we weighted a neighbor's impact on the colony of interest by its proximity to that colony.

Using the neighborhood definition above, we calculated the mean age of neighbors using an average weighted by distance from the colony of interest:

$$x_i = \frac{\sum_{j \in N_i} \frac{1}{d(i,j)} a_j}{\sum_{j \in N_i} \frac{1}{d(i,j)}}$$

Again, we performed grid search for the optimal number of neighbors k .

2.7 Voronoi Neighbors - Definition

We also consider an alternative definition of neighbors called Voronoi regions. This thinking about modeling ant colonies comes from [13].

For a colony i , we define its Voronoi region V_i across some space X with distance metric d (we use Euclidean distance) as follows:

Definition 2.2.

$$V_i = \{x \in X \mid d(x, V_i) \leq d(x, V_j) \forall j \neq i\}$$

In the below figure 3, we provide an example of a Voronoi partition for the ant population in question.

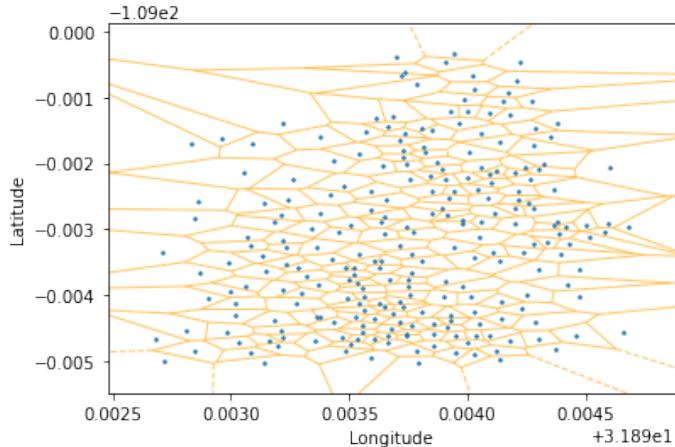


Figure 3: Example of Voronoi Partition: Ant Colonies in Rodeo, NM in 1989

Now, let R_Y denote the set of Voronoi ridges shared between colonies C_Y in year Y .

For each colony $i \in C_Y$, we will define the neighbors N_i of i as follows:

$$N_i = \{j \in C_Y \mid r_{i,j} \in R_Y\}$$

We investigated whether this model of neighborhoods may better reflect the true relationships and interactions between ant colonies.

2.8 Voronoi Neighbors - Uniform Interaction

First, we assume that each of Voronoi neighbors considered has the same level of interaction with the colony of interest.

Using the Voronoi definition of a neighborhood above, we calculated the mean age of neighbors as follows:

$$x_i = \frac{1}{|N_i|} \sum_{j \in N_i} a_j,$$

where a_j denotes the age of colony j .

We compute the leave-one-out CV balanced accuracy of the logistic regression model considering \mathbf{A} (as calculated in this section) and \mathbf{T} .

2.9 Voronoi Neighbors - Length-Weighted Interaction

Next, we weighted a neighbor's impact on the colony of interest by the size of their shared interaction frontier.

Using the previously defined Voronoi neighbors, we weighted our calculation of mean age of neighbors by the length of the ridge they shared with the colony of interest:

$$x_i = \frac{\sum_{j \in N_i} |r_{i,j}| a_j}{\sum_{j \in N_i} |r_{i,j}|}$$

Again, we computed the leave-one-out CV balanced accuracy of the logistic regression model considering \mathbf{A} (as calculated in this section) and \mathbf{T} .

2.10 Aggression Metric

Finally, we proposed a metric to estimate the aggression of a colony $i \in C$ for a year t .

Definition 2.3. $\forall j \in C$ s.t. $j \neq i$, let $I_{i,j,t'}$ denote whether i is the nearest neighbor of j in year t' and $d_{j,t'}$ denote whether j dies in year t' . We calculate the aggression metric $a_{i,t}$ as follows:

$$a_{i,t} = \sum_{j \in C} \sum_{t'} I_{i,j,t'} \cdot d_{j,t'}$$

This formula roughly represents the number of colonies which a given colony has been responsible for 'killing' so far.

We fit a logistic regression model with the aggression metric and age of the nearest neighbor, along with the average temperature in that year.

3 Results

3.1 Baseline Models

We present the CV-balanced-accuracy for the associated baseline models in table 1. From looking at this table, we can see that neighbor age, temperature, and humidity give the most indication of survival of a given colony. This is in comparison to metrics such as minimum distance (D) which is at the CV balanced accuracy threshold of 0.500 i.e. performing no better than random.

Variables	CV Balanced Accuracy
D	0.500
A	0.556
T	0.520
H	0.526
D, A	0.556
D, T	0.520
D, H	0.526
A, T	0.562
A, H	0.535
T, H	0.520
D, A, T	0.562
D, A, H	0.535
D, T, H	0.520
A, T, H	0.498
D, A, T, H	0.498

Table 1: CV Balanced Accuracy of Baseline Models with most significant accuracy results shown in bold. The threshold of CV balanced accuracy is 0.500.

3.2 Nearest Neighbors

The results of the unweighted and weighted nearest neighbors models are shown below in figures 4 and 5 respectively. From these, we see that in the unweighted version, one neighbor performs best while for the weighted version two neighbors provides a slightly higher leave-one-out-accuracy. This indicates that some young colonies may share meaningful and potentially detrimental interaction with more than one neighbor.



Figure 4: K Nearest Neighbors - Unweighted

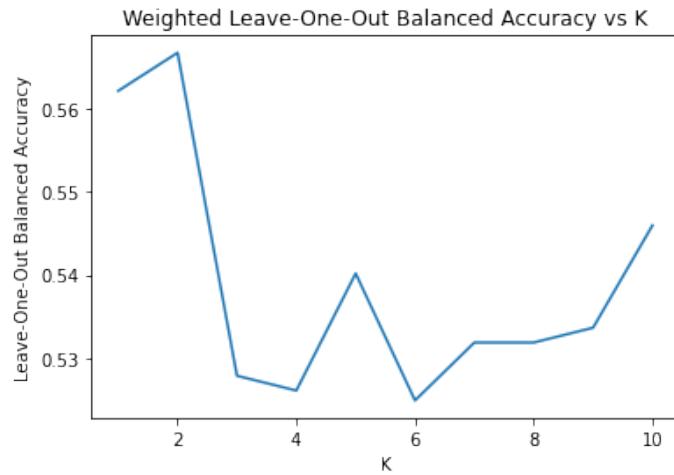


Figure 5: Nearest Neighbors - Weighted

3.3 Voronoi Partitions

For the Voronoi partition models, we simply present the associated accuracy metrics. These are given by model below:

(Uniform Weighted Interaction)

- **Mean Number of Neighbors:** 5.86
- **Leave-One-Out CV Balanced Accuracy:** 0.526

(Length-Weighted Interaction)

- **Leave-One-Out CV Balanced Accuracy:** 0.545
- **Average Percentile in Ridge Length of Nearest Neighbor:** 0.804
- **Percentage of Nearest Neighbors with Longest Ridge** 0.415

3.4 Aggression Metric

For the model which includes the proposed aggression metric as a predictor, we not only compute the leave-one-out CV balanced accuracy, but also the balanced accuracy on the test set, since it is the highest performing model:

- **Leave-One-Out CV Balanced Accuracy:** 0.723
- **Test Set Balanced Accuracy:** 0.870*

*No colonies die during the years in the test set, so this metric simplifies to standard, unbalanced accuracy.

4 Discussion

4.1 Tying Back into Biology

An initial point of interest as a result of the initial data analysis can be found in figures 6 and 7.

In figure 6, we see that as colonies age the range of observed distances to the nearest neighbor converges. This suggests that there is a stable range for this value at which a colony and their neighbors can effectively share resources. That is, if the colony is too far from its neighbors, it is likely positioned in location with less suitable resources for survival. Furthermore, if a colony is too close to its neighbors, it will experience heavy competition for the available resources.

In figure 7, we see that the age of the nearest neighbor colony decreases with age. This is primarily a function of the survivorship curve shown in 2. However, one notable feature of the plot in 7 are the thin, upward-trending bands, perpendicular to the aforementioned trend line. These bands suggest that certain pairs or groups of neighbors can co-exist, share resources, and “grow old together.”

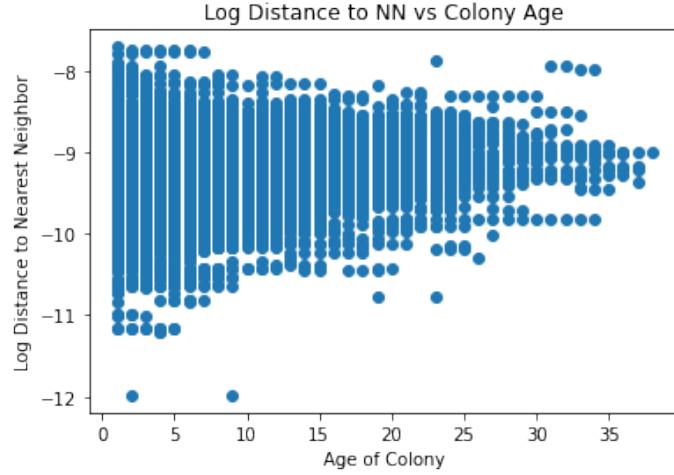


Figure 6: Log distance to nearest neighbor as a function of colony age

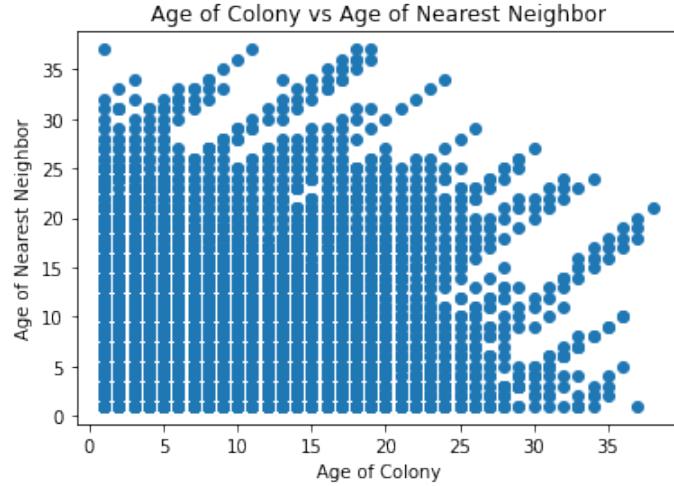


Figure 7: Age of nearest neighbor colony as a function of colony age

Looking back to our logistic regression models, we see several key takeaways. Firstly, the key factors to consider for prediction of one-to-two year life expectancy were temperature and age of nearest neighbor. This is confirmed by the grid search over k to show that 1-2 neighbors are the most impactful. Despite the results of [13], we found that distance was in fact a better predictor than other neighborhood definitions such as Voronoi partitions. This could simply be a result of more refined data processing in the literature versus our analysis. Secondly, in terms of aggression metrics, we see that the key takeaway

is that even a rudimentary aggression metric has strong predictive power for the one-to-two year life expectancy.

4.2 Future Work and Directions

Throughout our exposition of the results, we have attempted to flag points where assumptions have been made that could potentially be out of line with the biological state of the world. The most obvious next step is to incorporate the observations of [4] into the models. This can be done by obtaining individual foraging data and combining them both into the logistic regressions as an additional factor as well as investigating whether the evidenced effect of humidity on foraging [5] extends to calculations of aggression metrics.

The other takeaway of the provided results is to build both intuition and actionable information (colony ID lists) about where future field research efforts should be focused. This extends not only to future phenotyping, but also individual foraging and colony measuring. This kind of *post hoc* data analysis not only provides answers to previous questions but also allows the formulation of better questions in the future. Additionally, there are a wealth of field notes taken by researchers over the past 35+ years on the population in question. These can be combined with the quantitative results to either throw into question or verify various intuitions about aggressive neighbors and optimal number of neighbors for life expectancy.

References

- [1] Blair D. Beverly, H. McLendon, S. Nacu, S. Holmes, and D. M. Gordon. How site fidelity leads to individual differences in the foraging activity of harvester ants. *Behavioral Ecology*, 20(3):633–638, 2009.
- [2] Jacob D. Davidson, Roxana P. Arauco-Aliaga, Sam Crow, Deborah M. Gordon, and Mark S. Goldman. Effect of interactions between harvester ants on forager decisions. *Frontiers in Ecology and Evolution*, 4, Oct 2016.
- [3] Deborah Gordon. *Ant encounters:interaction networks and colony behavior*. Primers in complex systems. Princeton University Press, Princeton, NJ, course book edition, 2010.
- [4] Deborah M. Gordon. The rewards of restraint in the collective regulation of foraging by harvester ant colonies. *Nature*, 498(7452):91–93, Jun 2013.
- [5] Deborah M. Gordon, Katherine N. Dektar, and Noa Pinter-Wollman. Harvester ant colony variation in foraging activity and response to humidity. *PLoS ONE*, 8(5):e63363, May 2013.
- [6] Deborah M. Gordon, Adam Guetz, Michael J. Greene, and Susan Holmes. Colony variation in the collective regulation of foraging by harvester ants. *Behavioral Ecology*, 22(2):429–435, 2011.

- [7] Deborah M. Gordon and Alan W. Kulig. Founding, foraging, and fighting: Colony size and the spatial distribution of harvester ant nests. *Ecology*, 77(8):2393–2409, Dec 1996.
- [8] DeborahM. Gordon. How colony growth affects forager intrusion between neighboring harvester ant colonies. *Behavioral Ecology and Sociobiology*, 31(6), Dec 1992.
- [9] M. J. Greene and D. M. Gordon. Interaction rate informs harvester ant task decisions. *Behavioral Ecology*, 18(2):451–455, Mar 2007.
- [10] J. R. B. Lighton and George A. Bartholomew. Standard energy metabolism of a desert harvester ant, pogonomyrmex rugosus. effects of temperature, body mass, group size, and humidity. *Proceedings of the National Academy of Sciences*, 85(13):4765–4769, Jul 1988.
- [11] John R. B. Lighton and Donald H. Feener. Water-loss rate and cuticular permeability in foragers of the desert ant pogonomyrmex rugosus. *Physiological Zoology*, 62(6):1232–1256, Nov 1989.
- [12] Harold G. Nagel and Carl W. Rettenmeyer. Nuptial flights, reproductive behavior and colony founding of the western harvester ant, pogonomyrmex occidentalis (hymenoptera: Formicidae). *Journal of the Kansas Entomological Society*, 46(1):82–101, 1973.
- [13] Mekala Sundaram, Erik Steiner, and Deborah M. Gordon. Rainfall, neighbors, and foraging: The dynamics of a population of red harvester ant colonies 1988–2019. *Ecological Monographs*, 92(2):e1503, 2022.
- [14] Visual Cross Weather. Visual crossing weather. <https://www.visualcrossing.com/>, Visual Crossing Corporation, 1984–2022.