

Team Control Number

12600

Problem Chosen

A

2022

HiMCM

Summary Sheet

Animal pollinators have been supporting the lives of human beings on Earth. Bee pollinators are the biggest contributors to the pollination of crops, providing humans with food. Given such circumstances, this paper investigates the population of honeybee colonies and the processes of bee pollination.

We constructed Honeybee Colony Population Model (BCPM) to predict the population of a honeybee colony over time. We first outlined the life cycle of a honeybee, including eggs, larval stage, pupal stage, and adult bee stage. Within the adult bee stage, bees transition back and forth between foragers and hive bees depending on the number of available resources and the workload of nursing tasks. By listing out factors that affect the population in each stage, we established equations representing the rate of change in each of the stages of a honeybee's life cycle, as well as an equation describing the change in resource storage. We also evaluated the death rate and the resources in each month of the year and calculated each group's typical maximum, minimum, and mean population in a honeybee colony: 3 years after the establishment of the colony, the total adult population follows a seasonal change with recurring patterns each year, giving a maximum of 100862 bees and a minimum of 35676 bees. The annual average population is found to be 64877 bees.

We then conducted a sensitivity analysis on BCPM and found that the initial number of bee hives and the initial amount of available resources have the most significant impact on the population of the colony. We also observed an unusual pattern in the cross-analysis of the two factors and constructed Simplified Colony Collapse Disorder Model (SCCDM) to predict whether a colony will collapse using only one equation.

In response to question 3, Hive Deployment Model (HDM) is constructed to estimate the number of hives needed to support the pollination of a specific land area. We first divided the land into 20 nodes and then found the most appropriate locations to place the hives. After establishing the equations for movements between nodes per day per forager group, we developed an iterating algorithm to find the number of hives needed to pollinate crops on 20 acres of land. We collected data for 9 typical bee-pollinated plants and found the number of hives needed for each type of plant based on the algorithm, with blueberries being the most demanding, requiring 83 hives, whilst apples and roses only required 2 hives at the other end of the spectrum. Then, we established a sensitivity analysis to ensure the stability of the model by changing two arbitrary parameters.

Finally, we discussed the potential advantages and disadvantages of our model. We have also created a non-technical blog that summarizes our investigation, presenting our results in a simplified way.

Keywords: Honeybee, Colony Collapse Disorder, Least Squares Fitting, Differential Equations, Sensitivity Analysis, Undirected Weighted Graphs



To Bee Or Not to Bee, That is a Question

A BEEKEEPER'S GUIDE TO SUCCESS

In the biting winter of 2006, hundreds of beekeepers were heartbroken to find out that their prized beehives emptied overnight. Thousands of worker bees simply disappeared with no dead bodies in sight, leaving behind a lonely queen who had the slimmest chance of surviving the harsh cold by herself. Scientists later defined this distressing phenomenon as Colony Collapse Disorder (CCD), but they are still baffled by the reasons behind CCD to this day. As a beekeeper, I feel the horror of losing all of my bees near and deep, and bees are not exactly low-maintenance. This blog, therefore, is a guide to beekeeping distilled from my years of experience, theorized by my attempt at math modeling. If you are a fellow beekeeper or even just a bee enthusiast, read on!

To tackle the mystery of bee population, I racked my brain to answer the million-dollar question—which factor, exactly, has the biggest impact on honeybee colony size? Pages of equations and days of coding later, I identify two most important factors: the initial bee population and the number of initial resources, outcompeting the influence of forager number, egg number, larva number, etcetera. You may have noticed the emphasis on “initial.” Yes, my model suggests that a strong start is the key to bee success, providing a powerful base for future development. Another interesting finding I stumbled upon is that the bee population tends to be at the extremes. Bee hives either have a healthy, stable population of around 60,000, or they completely collapse. So this is your reminder to closely monitor your bees and to respond to changes immediately, because bee population is all or nothing.

In addition to the population of a single bee colony, I wanted to see the dynamics of beehives in action. Therefore, I calculated the number of beehives needed to pollinate a 20-acre land with different types of plants, organized in the table on the right. If you are curious about math models and want to see them in greater details, make sure to check out the paper attached below! But before I go, allow me to end this blog on a Shakespearean note. “Full merrily the humble-bee doth sing, Till he hath lost his honey and his sting.” May your humble-bees (or honey bees, in our case) never lose honey or sting!

Plants	Number of hives needed
Alfalfas	11
Almonds	21
Apples	2
Avocados	3
Blueberries	83
Lillies	9
Onions	5
Roses	10
Sunflowers	2

Contents

1	Introduction	4
1.1	Background	4
1.2	Problem Restatement	4
2	Assumptions and Justifications	5
3	Honeybee Colony Population Model (BCPM)	5
3.1	Variables	5
3.2	Typical Honeybee Lifecycle	5
3.3	Mathematical Modeling of Population	6
3.4	Parameter Determination	8
3.5	Typical Model Results	9
3.6	Sensitivity Analysis	10
3.7	Simplified Colony Collapse Disorder Model (SCCDM)	12
3.8	Conclusion	13
4	Hive Deployment Model (HDM)	14
4.1	Variables	14
4.2	Model Establishment	15
4.3	Evaluation and Results	18
4.4	Sensitivity Analysis	18
4.5	Conclusion	19
5	Strengths and Weaknesses	19
5.1	Strengths	19
5.2	Weaknesses	19
A	Appendix	20
A.1	BCPM Implementation in Python	20
A.2	HDM Implementation in Python	22
	References	24

1 Introduction

1.1 Background

Primarily consisting of the queen, drones, and worker bees, one beehive can contain tens of thousands of bees. Bees demonstrate impressive communicative abilities for a group of this size, with pheromones acting as an important social glue that binds the colony together. In particular, the Queen mandibular pheromone (QMP) released by the queen bee regulates the social behavior of the hive, including swarming and mating, insofar as to facilitate the inhibition of ovary development in worker bees [1]. Other communication tools, such as bee dances, are also a testament to bee colonies' complexity.

These intelligent creatures pollinate one-third of the human diet, feeding billions of people all over the world [2]. Nowadays, more than 90 crops are dependent on bee pollination, including apples, squash, broccoli, and many more [3]. In addition to humans, bee pollination also helps feed 80% of all the birds in the US. Bees' value also translates into economic profits [4]. In the US alone, bees contribute \$20 billion to the agricultural sector every year, forming an indispensable component of the global economy [5].

However, The status quo of bees is nothing short of grim—the population of bees peaked in 1947 and has been declining since; nearly one in every ten wild bee species face extinction [6]. Every year, two out of three beekeepers in the US lose 40% of their bee colonies due to compounded reasons, including diseases, climate change, and habitat destruction [7]. A study published in *Science* found that around 50% of the 109 bee species investigated had been lost, and the remaining species see a decline in their pollination ability. Therefore, it is vital that we, the direct beneficiaries of bees, develop a sound understanding of the pollination capabilities and population trends of bee colonies [8].

1.2 Problem Restatement

- We are asked to construct a mathematical model (Model 1) that estimates the population over time of one honeybee colony. Because bees go through three developmental stages — egg, larva, and pupa — we should take into consideration the time each stage takes to determine the rate at which a bee hive produces new bees. We should also take into consideration the initial amount of resources that the bees can access so that there is a starting point for future projections.
- We should conduct sensitivity analysis on our Model 1 and thus determine the change of which initial variables lead to the most significant change in the average population when the colony is in a stable stage.
- We should construct another mathematical model (Model 2) which estimates the minimum number of hives needed to support pollination in a 20-acre parcel of land. Various factors, such as pollination efficiency and changes in the number of flowers, should be taken into account, and data from Model 1, such as bee population, can be reused in this model. Then, we should test our Model 2 on some plant species to yield results.
- Based on the data we acquired in the previous questions, we should put together a straightforward blog that beekeepers or farmers could use as a potential guide. In the blog, we should point out the factors that have the most influence over the bee population and give our estimate of the approximate hive numbers needed to cover the 20-acre land.

2 Assumptions and Justifications

- **Assumption 1:** All eggs are hatched after 3 days they are laid. The larval stage lasts for 6 days, and the pupal stage lasts for 10 days.

Justification 1: Typically, an egg laid by the queen will be hatched in 3 days, the larval stage lasts for approximately 5.5-6.5 days, and the pupal stage lasts for approximately 7.5-14.5 days [9]. For the purpose of constructing a model, we simplify this premise by taking averages of the duration in the larval and the pupal stage, which will hardly have an impact on our results since the bee population is large.

- **Assumption 2:** 50% of foragers die every day if no resource is available.

Justification 2: To simulate the effect of an exponential decrease in foragers, we set a rate of decline in population each day when the storage of resources falls below 0. We do not expect an immediate collapse of the colony, nor do we want to trivialize the impact of this shortage. Therefore, we set the rate of death to 50% to imitate a temperate decline in the population but preserve the colony's chances of recovering.

- **Assumption 3:** The 20-acre farmland has a rectangular shape consisting of 20 smaller 1-acre squares arranged in a 4×5 pattern.

Justification 3: Since we do not have any information about the shape of the farmland, we assume that it is close to a typical rectangular farmland. For the purpose of our model, we divided the farmland into 20 nodes, each a 1-acre square, so that it is easier to calculate the distance between farmlands. The model can be altered slightly so that it can be applied to other shapes of farmlands since the process of HDM is the same.

- **Assumption 4:** For each forager, 25 visits are made to each acre every time. Besides, foragers are assigned in groups of 200 and operate together.

Justification 4: The number of visits and foragers is so huge that it will strongly affect our model and undermine the importance of other factors in our model due to time constraints. Considering the visits and forager behaviors as a group not only helps us solve the problem of time constraints but also has little impact on the final data we get from the model because we are simply expecting all the groups to operate at an average efficiency. The fluctuations of individuals will not influence the bees' overall behavior.

3 Honeybee Colony Population Model (BCPM)

3.1 Variables

The table of variables used in this model are shown in Table 3.1.

3.2 Typical Honeybee Lifecycle

To establish a model for the population of a honeybee colony, we first draw an outline of the life cycle of a typical honeybee. The queen lays eggs, which will be taken care of by the worker bees (hive bees), and the eggs are hatched within 3 days, followed by the larval stage which lasts for 6 days, and the pupal stage which lasts for 7-14 days. The bees then enter the adult stage [9] (Assumption 1, 2).

Symbol	Description
t	Time from start of model
n_H	Number of hive bees
n_E	Number of eggs
n_B	Number of larva and pupa
n_F	Number of foragers
R	Mass of available resources
c	Rate of resource collection
θ	Rate of resource consumption
k	Rate of forager death
r	Rate of transition to foraging
l	Maximum egg laying rate
ϕ	Maximum eclosion rate
$L(t)$	Egg laid per day
$E(t)$	Daily eclosion number
$F(t)$	Net forager transition rate

Table 3.1: Variables Used in BCPM

After the bees fully mature, they become worker bees, responsible for nursing tasks in the hives [9]. However, they can transition into foragers, focusing on resource collection and pollination. The reverse transition can also occur when there is more need for nursing than foraging [10].

Factors such as resource availability and beehive population can alter the development cycle, as shown in Figure 3.1.

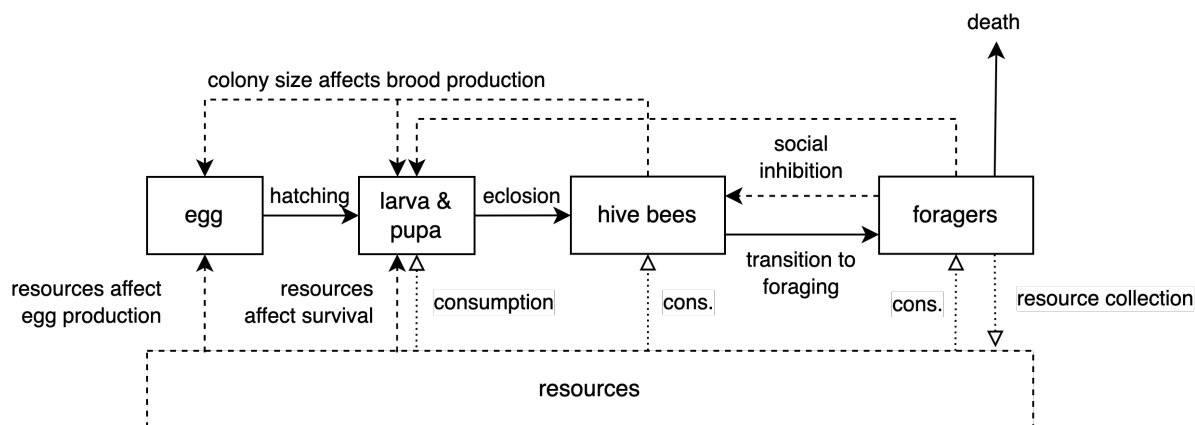


Figure 3.1: Graphical Representation of a Typical Honeybee Lifecycle

3.3 Mathematical Modeling of Population

We model the population trend by first considering the change in the population of each stage. Assuming that all eggs are hatched after 3 days they are laid (Assumption 1), we can derive:

$$\frac{dn_E}{dt} = L(t) - L(t - 3)$$

The laying rate per day $L(t)$ has a positive correlation to the number of available resources, but the effect reaches a maximum when the laying rate reaches a maximum for the queen. Similarly, the laying rate also positively correlates to the number of hive bees in the colony, but the effect reaches a maximum when there are excess workers. Therefore, we can construct the following equation:

$$L(t) = l \cdot \frac{R_t}{R_t + \varepsilon} \cdot \frac{n_{Ht}}{n_{Ht} + w}$$

where w and ε are arbitrary parameters that control the dependency of egg laying on the number of resources and the number of hive bees, respectively.

We then consider the rate of change of population in the larval and the pupal stages. The rate is equivalent to the hatching rate deducted by the eclosion rate, as presented in Figure 3.1. We estimate that the time taken between hatching and eclosion is 16 days (Assumption 2), which means that eclosion happens on the 19th day after an egg is laid. Thus, the following equation can be constructed:

$$\frac{d n_B}{d t} = L(t - 3) - L(t - 19)$$

where $L(t - 19)$ is derived from $L((t - 3) - 16)$.

Next, we try to model the rate of change of hive bees. The increase in the number of hive bees each day is equivalent to the eclosion rate from newly grown pupa subtracted by the net forager transition rate. The eclosion rate is dependent on the quantity of essential resources and the number of workers that take care of them during larval and pupal stages. We define the hive-bee population rate and eclosion rate as follows:

$$\frac{d n_H}{d t} = E(t) - F_R(t)$$

$$E(t) = \frac{R_t}{R_t + \alpha} \cdot L(t - 19) \cdot \frac{n_{Ft} + n_{Ht}}{n_{Ft} + n_{Ht} + \beta}$$

where the remaining population of $L(t - 19)$ are assumed to be dead. α and β are arbitrary parameters that determine the relevance of the factors of resources and population, respectively, on broods' survival.

The net forager transition rate is the rate of transition to foragers subtracted by the rate of inhibition, as shown in Figure 3.1. We represent this relationship in the following equation:

$$F_R(t) = n_{Ht} \cdot \left(r_{\min} + r \left(\frac{\gamma}{\gamma + R_t} \right) - \sigma \left(\frac{n_{Ft}}{n_{Ft} + n_{Ht}} \right) \right)$$

where γ and σ are arbitrary parameters that control the sensitivity of our model towards resources and forager-hive bees ratio, respectively. r_{\min} is the rate of forager recruitment when there are maximum resources but no foragers yet. This can be concluded when $R_t \rightarrow \infty$ and $n_{Ft} = 0$:

$$F_R(t) = n_{Ht} \cdot (r_{\min} + r(0) - \sigma(0)) = n_{Ht} \cdot r_{\min}$$

The rate of increase in the population of foragers is modelled as the net forager transition rate deducted by the death rate, which is shown as follows:

$$\frac{d n_F}{d t} = \begin{cases} n_{Ht} \cdot F_R(t) - k_r k_{\max} k_s n_{Ft} & \text{if } R_{t-1} + \Delta R_t \geq 0 \\ n_{Ht} \cdot F_R(t) - \frac{1}{2} n_{Ft} & \text{otherwise} \end{cases}$$

where k_r is a random real number between $[0.75, 1.25]$.

The first condition is satisfied when there are still resources to meet the basic survival of foragers, so there is a natural death rate of $k_r k_{\max} k_s$. k_r simulates the random factors the foragers may encounter which affect their lifespan, fluctuating the daily death rate by $\pm 25\%$. $k_s k_{\max}$ is the seasonality factor, which determines the death rate based on typical patterns observed throughout the year, and is further explained in Section 3.3.

When no resource is available, we set the forager death rate to 0.5 to simulate an exponential decline in population.

Lastly, we establish an equation to describe the change in resources over time. The resource is collected by foragers but consumed by larvae, pupa, hive bees, and foragers. Therefore:

$$\frac{d R}{d t} = c_r c_{\max} c_s n_{Ft} - \theta_B n_{Bt} - \theta_H n_{Ht} - \theta_F n_{Ft}$$

where c_r is a random real number between $[0.75, 1.25]$.

Similar to the seasonality factor for death rate, $c_s c_{\max}$ is the rate of resource collection subject to seasonal change. c_r is used to simulate the fluctuating resource collection by $\pm 25\%$ due to unseen random factors, which is further described in Section 3.3.

3.4 Parameter Determination

We set the maximum laying rate of the queen $l = 2000$, in accordance with real-world observations and research [11]. According to Harbo [12], a brood typically requires 163 mg of resources to develop into an adult, which is equivalent to 10.2 mg per day over 16 days; an adult bee requires 6.7 mg resources per day. Thus, we take $\theta_B = 0.0102$ and $\theta_F = \theta_H = 0.0067$.

We set $r_{\min} = 0.25$ and $\sigma = 0.75$ to assume that hive bees will only transform into foragers 4 days after their eclosion, and set $r = 0.25$ to double forager recruitment when there is a shortage of resources, which is consistent with Khoury et al.'s suggestions [13].

We also set $w = 5000$ to represent the effect that half of the eggs fail to hatch when only 5000 hive bees take care of them. Then, we set $\alpha = \gamma = \varepsilon = 1000$ to show the decreased effect when total resources stored exceeds 1 kg, as $\frac{R_t}{R_t + \alpha} > \frac{1}{2}$, $\frac{R_t}{R_t + \varepsilon} > \frac{1}{2}$, and $\frac{\gamma}{\gamma + R_t} < \frac{1}{2}$ when $R_t > 1000$.

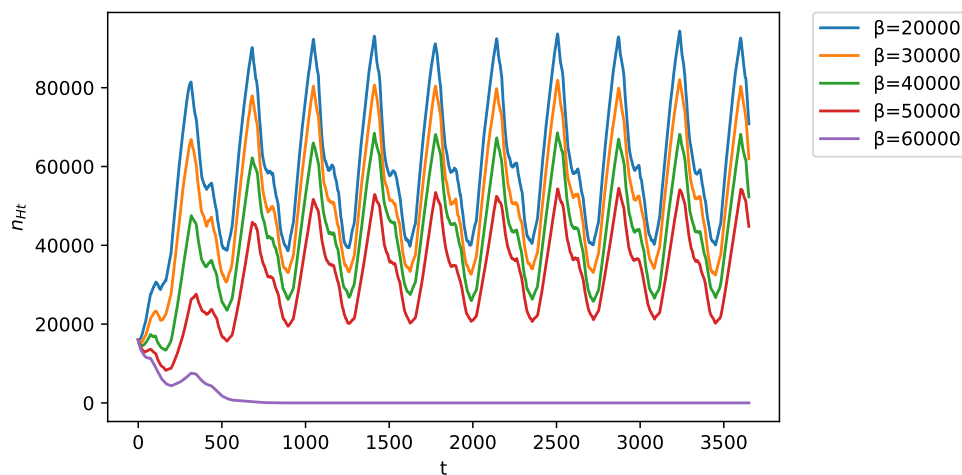
According to Russell et al. [14], the death rate k of foragers and the resource collection rate c is normalised to give maximum values $k_{\max} = 0.102$, $c_{\max} = 0.099$ and rates k_s , c_s per month, as shown in Table 3.2.

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
k_s	0.29	0.29	0.54	0.79	0.68	0.57	0.78	1.00	0.83	0.67	0.29	0.29
c_s	0.03	0.03	0.48	0.93	0.97	1.00	0.60	0.19	0.32	0.45	0.25	0.10

Table 3.2: Table of k_s and c_s values

Lastly, we conducted several tests using initial values introduced in Section 3.4 on our model to determine a reasonable β value. According to the results shown in Table 3.3 and Figure 3.2, since a typical honeybee hive contains between 20,000 and 80,000 honeybees, we observe that $\beta = 40000$ fits our need for this model.

β Value	Maximum n_{Ht} ($t > 1000$)	Minimum n_{Ht} ($t > 1000$)
20000	93635	38473
30000	82591	32149
40000	69531	25930
50000	54743	19674
60000	1	0

Table 3.3: Table of BCPM Results for Various β Values over 10 YearsFigure 3.2: BCPM Results for Various β Values Plotted over 10 Years

3.5 Typical Model Results

We define several initial values that represent a typical start-up phase of a honeybee colony. Initial eggs, larva and pupa number $n_{E0} = n_{B0} = 0$ since the colony has not yet created any brood. Initial hive bee number $n_{H0} = 16000$ and initial forager number $n_{F0} = 8000$, which add up to give a population slightly over the minimum of 20,000 and in an approximate ratio of 2 : 1. Initial resource store $R_0 = 5000$ to imitate the situation of basic storage at start-up.

We also set the starting as the 90th day of the year (in spring), when there are enough resources to support the colony's survival and growth, and this is the typical period when broods start to form and develop [15]. The model is evaluated for 10 years ($0 \leq t \leq 3650$) to

give the results in Table 3.4 and Figure 3.3. All maximum, minimum and average values are calculated when $t > 1000$.

Name	Symbol	Maximum	Minimum	Average
Total (adult) population	n	100862	35676	64877
Hive bee count	n_{Ht}	69459	25277	45313
Forager count	n_{Ft}	31737	10247	19622
Egg count	n_{Et}	1780	1190	1544
Larva and pupa count	n_{Bt}	9468	6392	8234

Table 3.4: Table of Typical BCPM Results

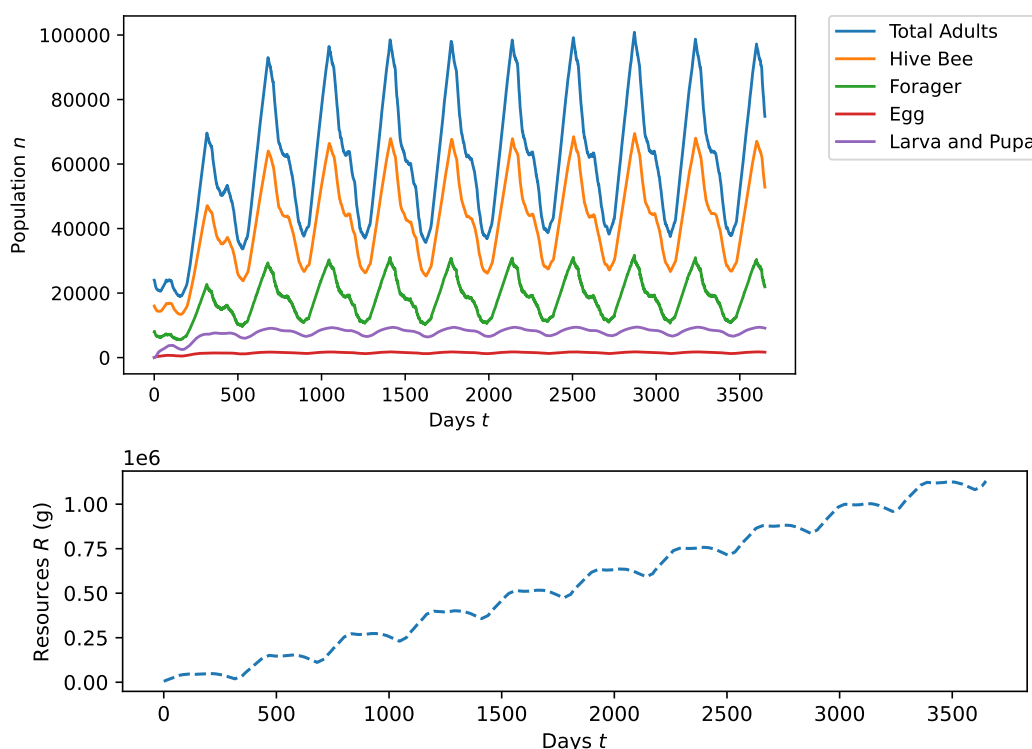


Figure 3.3: Typical BCPM Results

We observe a cyclical fluctuation in population and resource growth that occurs every year, which is reasonable due to the periodic seasonal death rate $k_{\max}k_s$ and resource collection rate $c_{\max}c_s$.

3.6 Sensitivity Analysis

3.6.1 Sensitivity Analysis on Arbitrary Parameters

Firstly, we change the arbitrary parameters r_{\min} , r , σ , α , γ and ε (the sensitivity analysis for β has been conducted in Section 3.3) by $\pm 10\%$ to test the consistency of the repeating pattern shown by our model. The results are shown in Figure 3.4 and 3.5.

We observe a generally consistent trend in all tests, so we consider our choices of parameters to be reasonable and stable overall.

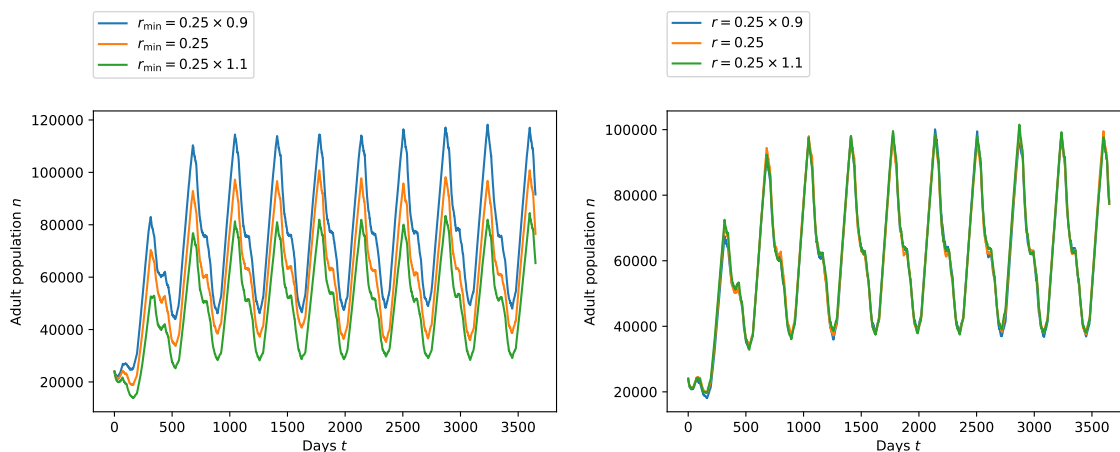


Figure 3.4: Sensitivity Analysis on Arbitrary Parameters

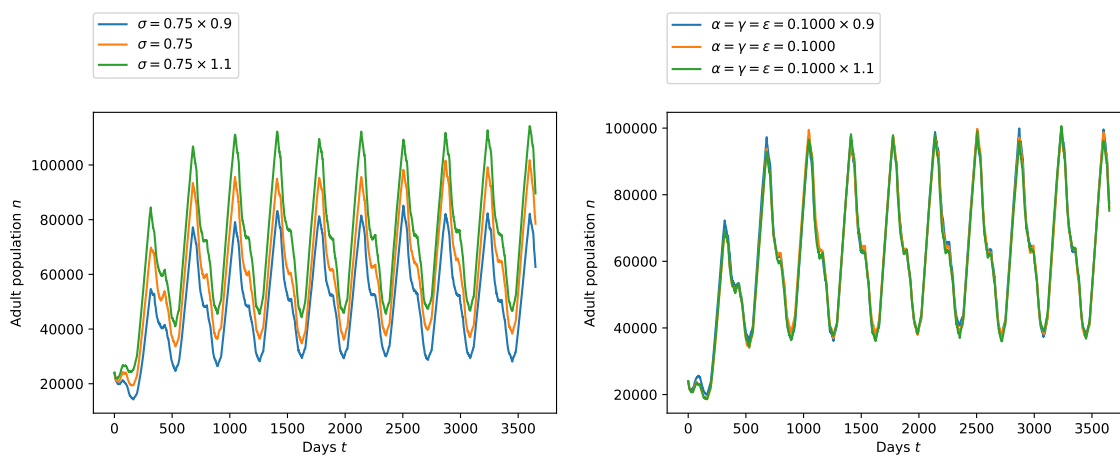


Figure 3.5: (cont.) Sensitivity Analysis on Arbitrary Parameters

3.6.2 Sensitivity Analysis on Initial Values

In order to identify the factors that have the most significant effect on the honeybee colony population, we conduct multiple simulations using our model while tweaking the initial values. The results are shown in Table 3.5.

We notice that the initial resource stored R_0 and initial hive bee population n_{H0} significantly affect the population compared to other factors: the honeybee population shrinks and eventually collapses when the values of these two factors decrease. Thus, we perform another sensitivity analysis based on both factors changing simultaneously, where $0 \leq R_0 \leq 10000$ and $0 \leq n_{H0} \leq 32000$ for $200^2 = 40000$ possible combinations for the two factors. We eliminated the effect of random terms k_r and c_r to reach a smoother graph in Figure 3.6.

We observe an unusual pattern– the honeybee population either maintains a relatively stable average number of around 60000 or the colony collapses completely. Therefore, we now construct an extra model which determines whether a colony collapses.

Variable	n			Variable	n		
R_0	max.	min.	mean	n_{F0}	max.	min.	mean
10000	68565	25398	45507	(4000)	69080	25662	45224
5000	68799	25889	45466	2000	68690	26500	45603
2500	68808	25953	45233	0	68629	26101	45265
1250	68949	24928	45166	n_{E0}	max.	min.	mean
0	—	—	[272 d]	32000	69179	25525	45425
n_{H0}	max.	min.	mean	16000	69385	26039	45528
32000	69052	25621	45392	8000	68482	25951	45372
16000	68607	25147	45188	4000	70005	26233	45690
8000	68411	26394	45379	0	68494	25369	44968
4000	68285	23064	43080	n_{B0}	max.	min.	mean
0	—	—	[117 d]	32000	69529	25961	45344
n_{F0}	max.	min.	mean	16000	68579	26525	45705
16000	69594	26296	46030	8000	68764	26486	45558
8000	68529	26143	45534	4000	68838	25863	45257
4000	69080	25662	45224	0	68698	25544	45243

*"—" denotes a collapsed colony; number in square brackets is the time the colony survived.

Table 3.5: Table of BCPM Results under Different Initial Values

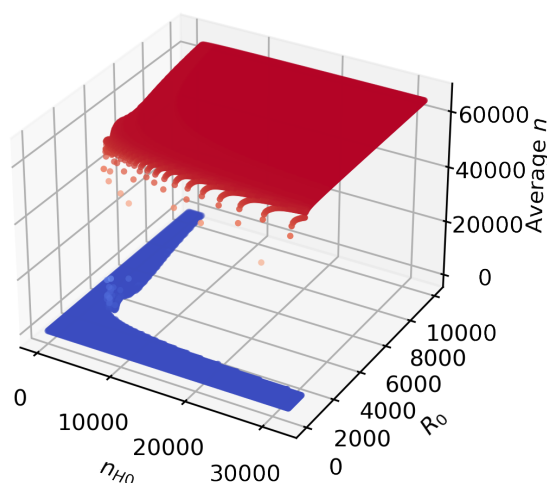


Figure 3.6: Sensitivity Analysis on n_{H0} and R_0

3.7 Simplified Colony Collapse Disorder Model (SCCDM)

We first set colonies with an average population below the threshold of $n = 30000$ to collapse; otherwise, the average population is set to 60000. Dimension reduction is also performed to fit the relationship into a planar equation. The process is shown in Figure 3.7.

We now trace the edge shown in Figure 3.7 (right) by selecting the minimum red n_{H0} value

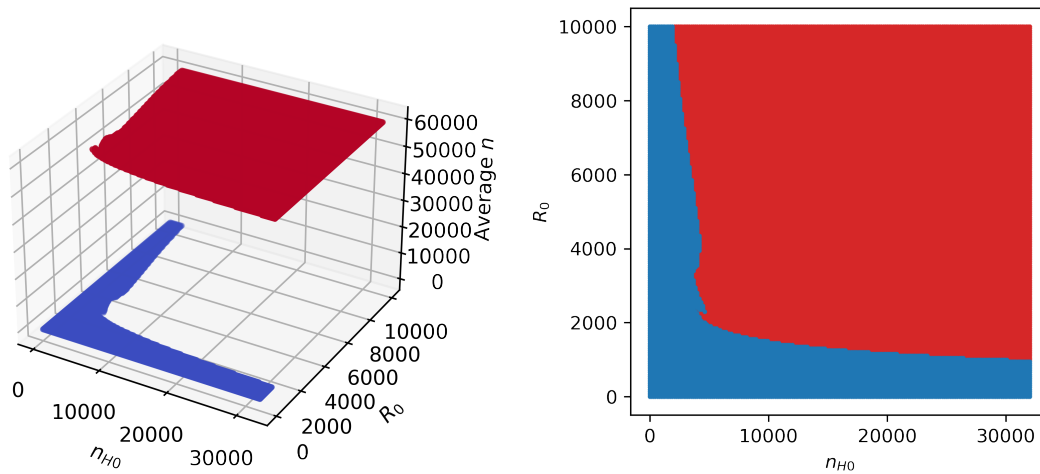


Figure 3.7: Thresholding and Dimension Reduction on Figure 3.6

for each R_0 . We can then fit the following equation to the curve using non-linear least squares:

$$R_0 = ae^{-bn_{H0}} + c$$

where a , b and c are constants to be determined. The results are plotted in Figure 3.8.

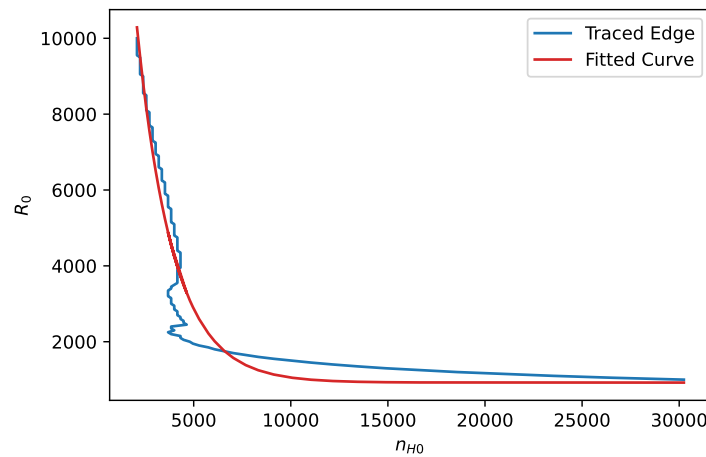


Figure 3.8: Fitted Function for SCCDM

Therefore, the evaluated constants, together with the rearranged fitted equation, state that a honeybee colony will only thrive when

$$R_0 - a \exp(-bn_{H0}) > c$$

where $a = 2.86660147 \times 10^4$, $b = 5.3785166 \times 10^{-4}$ and $c = 9.224787 \times 10^2$.

3.8 Conclusion

We have established a Honeybee Colony Population Model (BCPM) with a group of equations, as described in Section 3.2, to simulate the change in the honeybee population over time.

The typical results described in Table 3.4 and Figure 3.3 show a pattern of annual periodic change, which is reasonable due to the presence of seasonality factors k_s and c_s .

We have performed sensitivity analysis on all of our arbitrary parameters to reach a conclusion that our model is relatively stable and is hence reliable. We repeated the analysis for all initial values and found that the initial hive bee number n_{H0} and initial resource store R_0 are factors that are most significant in affecting the honeybee colony's population.

Lastly, we reached the conclusion that a colony either collapses or maintains a relatively constant average population at around 60000. The determining equation for colony survival is $R_0 - a \exp(-bn_{H0}) > c$ when $a = 2.86660147 \times 10^4$, $b = 5.3785166 \times 10^{-4}$ and $c = 9.224787 \times 10^2$.

4 Hive Deployment Model (HDM)

4.1 Variables

The table of variables used in this model is shown in Table 4.1.

Symbol	Description
\mathbb{H}	set of all hive locations
k_0	Number of flowers per acre
k	Number of unpollinated flowers
n_{Ft}	n_{Ft} from Model 1
c_s	c_s from Model 1
v	Maximum daily forager flower visits
m_{xy}	Simulated node for 1 acre
g	Number of foragers per group
p	Visits per forager per area
H	Number of hives deployed
ζ	Sum of minimum distances to nearest hive
f	Number of movement between nodes per day
d	Number of forager groups
a	Tiredness index
ω	Daily sum of costs taken by a group
s	Distance attenuation
p_r	Pollinations per group per area
r	Random negative effect
t_0	Starting day of pollination
t_s	Starting date of pollination in year
Δt	Duration of pollination
q	Population increment factor when $H > 20$

Table 4.1: Variables Used in HDM

4.2 Model Establishment

We first attempt to establish the model through a simulation approach. Firstly, we model the 20-acre farmland into 20 nodes on an undirected, weighted graph, as shown in Figure 4.1. The number marked on each node is the remaining flowers waiting to be pollinated, and the number on each edge marks the cost. We denote the bottom-leftmost node as m_{00} , and the upper-rightmost node as m_{34} .

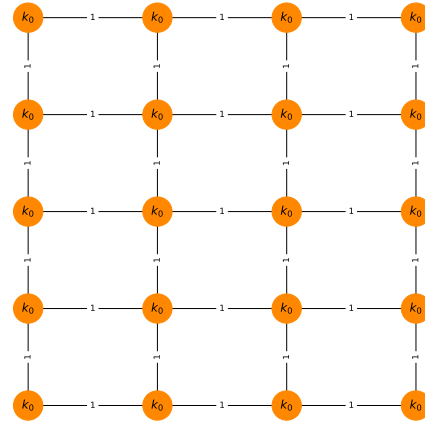


Figure 4.1: Initial Construction of Graph

The model is evaluated multiple times for hive number $H \geq 1$ until the number of hives can satisfy the pollination requirements. For each choice of H , we consider the "best" locations to place the hives so that the sum of distances from each node to their respective nearest hive ζ is minimal. This relationship is expressed as the following equation:

$$\zeta = \sum_{x=0}^3 \sum_{y=0}^4 \min\{|x - x_h| + |y - y_h| \mid (x_h, y_h) \in \mathbb{H}\}$$

where \mathbb{H} is the set of all hive locations. For $1 \leq H \leq 6$, we calculate ζ for all possible $\binom{20}{1} + \binom{20}{2} + \dots + \binom{20}{6} = 60459$ permutations to find out the best arrangement of the hives, as shown in Figure 4.2, where all nodes are represented in the structure same as Figure 4.1, and a blue circle indicates the presence of a hive. For $7 \leq H \leq 20$, an extra hive is put in a random location without a hive according to the arrangement of $H - 1$, since ζ reaches $20 - H$ in this range, and every extra hive only contributes a decrease of 1 in ζ due to the decrease in distance of a current node from 1 to 0.

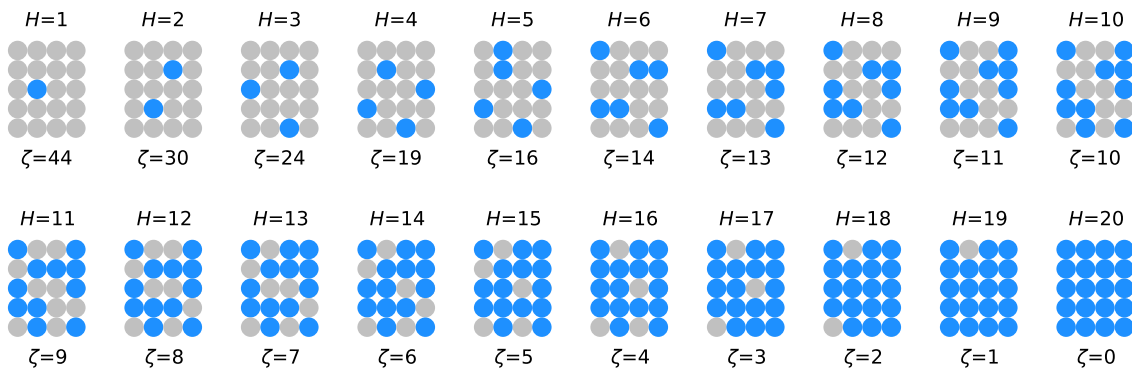


Figure 4.2: Hive Arrangements for Various ζ Values

In order to reduce the time complexity of our model, we assume that $p = 25$ visits are made to each node each time. We also group the foragers into groups of $g = 200$ in which group members operate together. We find the total number of movements between nodes per day

$f = \lfloor v/p \rfloor$, where v is the maximum flower visits per forager per day. Therefore, for each day $t \in [t_0, t_0 + \Delta t]$ where $t_0 = 365 \times 3 + t_s$, we consider the followings for each hive at m_{x_h, y_h} :

1. $d = \lfloor n_{Ft}/g \rfloor$ groups of foragers are formed and are all placed at m_{x_h, y_h} .
2. the number of flowers that require pollination increases by $\frac{1}{50}(k_0 - k)$, assuming that over 80% of flowers will require pollination again in 90 days if all flowers are pollinated.
3. For each group of foragers, all costs (weights on edges) are reset to 1 and steps 3-5 are repeated for $f c_s$ times. c_s is the seasonality factor used in BCPM for simulating changing resource collection rates and is therefore used to represent the change in pollination due to seasonal changes such as temperature and weather.
4. The group pollinates $p_r = g \cdot \frac{k}{k_0} \cdot p \cdot a \cdot s \cdot r$ flowers in current node, where
 - $\frac{k}{k_0}$ is the possibility that a randomly chosen flower is not pollinated. This factor shows the effect that a honeybee is more likely to visit a pollinated flower if the proportion of pollinated flowers is already large, decreasing successful pollination rates.
 - The tiredness index $a = \frac{v}{v+\omega}$ is used to address the effect of falling pollination rate due to working strain during a single day.
 - The distance attenuation $s = \frac{1}{1+s_0^2}$ shows halved pollination rate when the foragers are operating at neighbouring nodes from the hive, and gives an increasing effect as the distance increases. Since bees have a preference to work within 300 feet from their hive [16], we consider this factor as reasonable as neighbouring nodes are at a distance of approximately 209 to 418 feet from the centre node (considering each node is square farmlands; each farmland has side lengths of $\sqrt{43560 \text{ ft}^2} \approx 209 \text{ ft}$).
 - The random negative effect r is used to simulate random effects such as storms and diseases in crops. We take this effect as an arbitrary random number where $r \in [0.75, 1]$.
5. Then, the group transfers to another neighbouring node, such that the cost to travel to the node is minimum. If there are multiple nodes that satisfy this condition, then a random node between them is chosen.
6. The cost of the path that the group has just passed increases by 1. This reduces the likelihood of the group repeating the same paths in a short time.

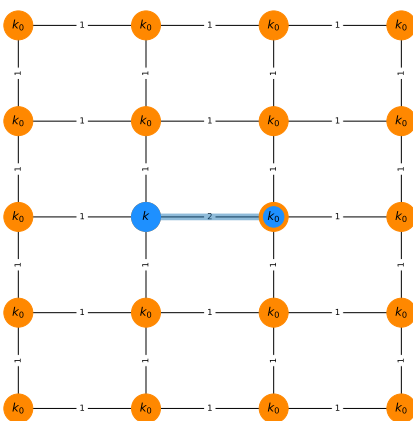


Figure 4.3 shows a typical pattern after simulating steps 1-6 from the initial graph in Figure 4.1, without any repetition involved. The blue circle indicates the hive location, the blue lines indicate the path the group has just passed, and the blue-in-orange circles indicate the resulting location of the group.

Following this algorithm, the number of yet-to-be-pollinated flowers on each node can be found for one specific number of hives. We repeat this algorithm with an increasing number of hives until we reach $< 1\%$ of

Figure 4.3: Graph Pattern after First Movement

total not pollinated crops in the 20-acre area, and $< 5\%$ of not pollinated crops in each individual node. If these conditions are not satisfied even when $H = 20$, we multiply the population in each hive by a factor q until the conditions are met, where we would conclude that $H = 20q$ hives are needed.

A pseudo code representation of this model is shown in Algorithm 1.

Algorithm 1 Simulation Process of HDM

```

 $G \leftarrow$  weighted undirected graph in Figure 4.1
 $H, h \leftarrow 1$ 
while  $\max k \geq 0.05k_0$  or  $\Sigma k \geq 0.01 \times 20k_0$  do
  if  $H > 20$  then
     $h \leftarrow 20$ 
     $q \leftarrow H/20$ 
  end if
  for all  $t \in [t_0, t_0 + \Delta t]$  do ▷ for each day
     $d \leftarrow \lfloor qn_{Ft}/g \rfloor$ 
    all  $k \leftarrow (k - k_0)/50$  ▷ add new flowers each day
    for all  $(x_h, y_h) \in \mathbb{H}_h$  do ▷ for each hive
      for  $1 \dots d$  do ▷ for each group
        All weights on graph  $\leftarrow 1$ 
         $\omega \leftarrow 0$ 
         $x, y \leftarrow x_h, y_h$ 
        for  $1 \dots \lfloor fc_{st} \rfloor$  do ▷ for each visit
           $a \leftarrow \frac{v}{v+\omega}$ 
           $s \leftarrow \frac{1}{1+s_0^2}$ 
           $k \leftarrow k - g \cdot \frac{k}{k_0} \cdot p \cdot a \cdot s \cdot \text{RANDOM}(0.75, 1)$  ▷ pollination occurs
           $\omega \leftarrow \omega +$  next node's cost
          Move to next node and cost + 1
        end for
      end for
    end for
     $H \leftarrow H + 1$ 
  end while

```

As an example, for arbitrary parameters $H = 1, v = 600, t_s = 90, \Delta t = 60$ and $k_0 = 150000$, we obtain a reasonable distribution of remaining not pollinated flowers, where the number increases as the distance from hive $(1, 2)$ increases. This typical result is shown in Figure 4.4.

In this stage, the hive is not considered to support pollination of the farmland, since the maximum k at $(3, 4)$ is 35093, which is a number far larger than $150000 \times 0.05 = 7500$. Moreover, the total number of not pollinated flowers is 225482, and it is still far larger than the require-

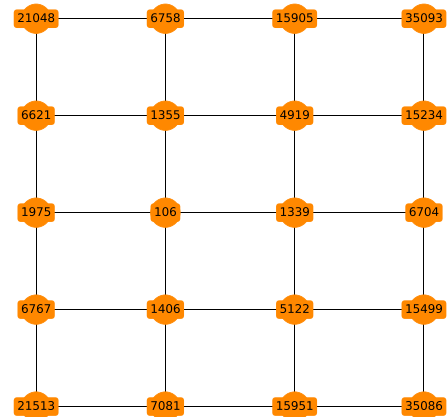


Figure 4.4: Typical Result of HDM

ment $150000 \times 20 \times 0.01 = 30000$.

4.3 Evaluation and Results

We first set the maximum daily flower visits per forager $v = 600$, which is a crop species-inspecific parameter and is held constant across different evaluations of our model. Since a bee colony can pollinate 20 million flowers per day at maximum [17], we divide the number by the typical daily maximum forager count $n_{F \max} \approx 30000$ in Table 3.4. Thus, it gives individual maximum flower visits $v = 2 \times 10^7 \div 30000 = 600$.

Then, we choose 9 widely seen bee-pollinated plant species to consider their respective s , Δt and k_0 . The results are shown in Table 4.2.

Species	k_0	s [18]	Δt [18]	Hives needed	Hives / acre
Alfalfas	266667	165	40	11	0.55
Almonds	860000	30	45	21	1.05
Apples	14400	135	9	2	0.10
Avocados	32000	90	60	3	0.15
Blueberries	3333333	97	49	83	4.15
Lilies	180000	157	35	9	0.45
Onion	80000	120	30	5	0.25
Roses	150000	150	90	10	0.50
Sunflowers	25000	200	84	2	0.10

Table 4.2: HDM Results for Various Species

4.4 Sensitivity Analysis

Lastly, we perform sensitivity analysis to ensure the stability and reliability of our model's outcomes. We alter the arbitrary variables r and daily increase in not pollinated flowers by:

- Changing the lower bound for r by $\pm 10\%$, so that $r \in [0.75 \pm 0.075, 1]$.
- Changing the factor of daily increase in not pollinated flowers by $\pm 10\%$, so that the increase becomes $(\frac{1}{50} \pm \frac{1}{500})(k_0 - k)$.

We randomly change the two variables in the $\pm 10\%$ range for 50 random combinations, then test and compare the outcomes on lillies, whose $k_0 = 180000$, $s = 157$, $\Delta t = 35$ and require 10 hives according to previous results in Table 4.2. The tested results are plotted as shown in Figure 4.5.

We notice that all tests evaluate an outcome of $H = 9$. Therefore, we consider the results to be stable and less prone to the arbitrary parameters mentioned above.

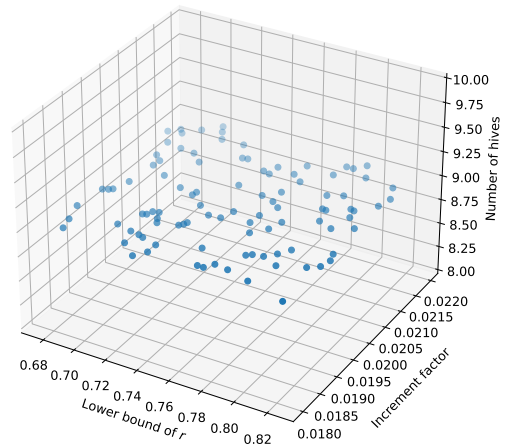


Figure 4.5: Sensitivity Analysis of HDM

4.5 Conclusion

We have established a model that estimates the number of hives needed in the 20-acre field for different plant species depending on flower density, pollination date, and pollination duration. The model uses a simulation algorithm that takes each acre as a 1-acre node and imitates the operation of groups of foragers.

We have given some results for 9 widely grown, bee-pollinated plant species. Among these species, blueberries take a maximum of 83 hives (4.15 hives / acre), whereas apples and sunflowers take a minimum of 2 hives (0.1 hive / acre).

Lastly, we conducted a sensitivity analysis on arbitrary parameters in our model. The analysis reported a consistent result through $\pm 10\%$ changes in those parameters. Therefore, we consider our model to be stable.

5 Strengths and Weaknesses

5.1 Strengths

- Honeybee Colony Population Model (BCPM) estimates the population of a typical colony, including hive bees, foragers, eggs, larvae, and pupae. This model not only estimates the overall population over time, but also gives insight into other aspects, such as the adult proportion in the colony or egg laying rate of the queen.
- We proposed a simplified equation in Simplified Colony Collapse Disorder Model (SC-CDM) to predict whether or not a colony will collapse, depending on its initial hive bee number and resource store. Therefore, the model is relatively easy to implement and run with faster evaluation.
- The time complexity of the Hive Deployment Model (HDM) has been optimized, which is around 5000 times faster with the optimization. The typical time of evaluation used for each species is around 15 seconds, which is considered acceptable as this model is typically run for only one species at a time.
- HDM also suggests the best locations to place hives which maximizes the pollinated flowers with minimal hive numbers. Minimizing the number of hives used reduces the cost and effort required when deploying these hives in real-life scenarios.

5.2 Weaknesses

- We did not consider the effect of unexpected events such as diseases and natural disasters in BCPM. Since these events heavily depend on locations and background, this effect is hardly predictable with the given information. However, since the occurrence of these events is extremely rare, their influence on the overall results in our model is not significant.
- Considering the complexity of HDM, we did not evaluate the negative impacts over time in detail, instead, we used a random negative effect coefficient r to demonstrate the overall effect, whose effect may be less reliable but has the advantage of reducing the complexity of our model.

A Appendix

A.1 BCPM Implementation in Python

```

import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

c_max = 0.099
c_season = np.array([0.03] * 31 + [0.03] * 28 + [0.48] * 31 +
                    [0.93] * 30 + [0.97] * 31 + [1.00] * 30 +
                    [0.60] * 31 + [0.19] * 31 + [0.32] * 30 +
                    [0.45] * 31 + [0.25] * 30 + [0.10] * 31)

k_max = 0.102
k_season = np.array([0.29] * 31 + [0.29] * 28 + [0.54] * 31 +
                    [0.79] * 30 + [0.68] * 31 + [0.57] * 30 +
                    [0.78] * 31 + [1.00] * 31 + [0.83] * 30 +
                    [0.57] * 31 + [0.29] * 30 + [0.29] * 31)

l = 2000
sigma = 0.75
r_min = 0.25
r = 0.25
alpha = gamma = epsilon = 1000
beta = 40000
w = 5000
theta_b = 0.0102
theta_f = theta_h = 0.0067

n_hive_0 = 16000
n_forager_0 = 8000
resources_0 = 5000
n_eggs_0 = 0
n_larva_0 = 0
n_hive = np.array([n_hive_0] * 19, dtype=float)
n_forager = np.array([n_forager_0] * 19, dtype=float)
resources = np.array([resources_0] * 19, dtype=float)
n_eggs = np.array([n_eggs_0] * 19, dtype=float)
n_larva = np.array([n_larva_0] * 19, dtype=float)

# day 0
n_hive = np.append(n_hive, n_hive_0)
n_forager = np.append(n_forager, n_forager_0)
resources = np.append(resources, resources_0)
n_eggs = np.append(n_eggs, 0)
n_larva = np.append(n_larva, 0)

def L(t):
    return l * resources[t] / (resources[t] +

```

$$\text{epsilon}) * n_hive[t] / (n_hive[t] + w)$$

```

def E(t):
    return resources[t] / (resources[t] + alpha) * L(t - 19) * (
        n_forager[t] + n_hive[t]) / (n_forager[t] + n_hive[t] + beta)

def F(t):
    return r_min + r * (gamma / (gamma + resources[t])) - sigma * (
        n_forager[t] / (n_forager[t] + n_hive[t]))

t_offset = 90
RANDOM = True
for t in range(19, 19 + 365 * 10):
    if RANDOM:
        diff_resources = (
            c_max * c_season[(t + t_offset) % 365] *
            (0.75 + np.random.random(1) / 2) *
            n_forager[t]) - theta_b * n_larva[t] - theta_h * n_hive[
                t] - theta_f * n_forager[t]
    else:
        diff_resources = (c_max * c_season[
            (t + t_offset) % 365] * n_forager[t]) - theta_b * n_larva[
                t] - theta_h * n_hive[t] - theta_f * n_forager[t]
    resources = np.append(
        resources, np.maximum(resources[t] + diff_resources, 0))
    diff_egg = L(t) - L(t - 3)
    n_eggs = np.append(n_eggs, np.maximum(n_eggs[t] + diff_egg, 0))
    diff_larva = L(t - 3) - L(t - 19)
    n_larva = np.append(n_larva, np.maximum(n_larva[t] + diff_larva,
        0))
    diff_hive = E(t) - n_hive[t] * F(t)
    n_hive = np.append(n_hive, np.maximum(n_hive[t] + diff_hive, 0))
    if (resources[t] + diff_resources) < 0:
        k = 0.5
    else:
        k = k_max * k_season[(t + t_offset) % 365]
    if RANDOM:
        diff_forager = n_hive[t] * F(t) - k * n_forager[t] * (
            0.75 + np.random.random(1) / 2)
    else:
        diff_forager = n_hive[t] * F(t) - k * n_forager[t]
    n_forager = np.append(n_forager,
        np.maximum(n_forager[t] + diff_forager, 0))

```

A.2 HDM Implementation in Python

```

import numpy as np
import networkx as nx
import pandas as pd

# hive loc is hidden due to length constraints.
c_season = np.array([0.03] * 31 + [0.03] * 28 + [0.48] * 31 +
                    [0.93] * 30 + [0.97] * 31 + [1.00] * 30 +
                    [0.60] * 31 + [0.19] * 31 + [0.32] * 30 +
                    [0.45] * 31 + [0.25] * 30 + [0.10] * 31,
                    dtype=np.float32)

paths = dict(nx.all_pairs_dijkstra(G))
n_forager = np.array(pd.read_csv('n_forager.csv')['0'])
def model2(t_start=90,
          t_delta=60,
          k_flower=200000,
          hive_count=1,
          r_lower=0.75,
          inc=1 / 50):
    print("AIM", 0.01 * k_flower * 20, 0.05 * k_flower)
    group_size = 200
    p_pol = 25
    v_visit = 600
    q_multiply = 1
    G = nx.Graph()
    for x in range(4):
        for y in range(5):
            G.add_node((x, y))
    for x in range(4):
        for y in range(5):
            if x < 3:
                G.add_edge((x, y), (x + 1, y), weight=1)
            if y < 4:
                G.add_edge((x, y), (x, y + 1), weight=1)
    nx.set_node_attributes(G, k_flower, 'flower')
    while True:
        if hive_count > 20:
            hives = hive_loc[20]
            q_multiply = hive_count / 20
        else:
            hives = hive_loc[hive_count]
            q_multiply = 1
        for t in range(t_start, t_start + t_delta + 1):
            f_reps = v_visit // p_pol
            d_num = q_multiply * n_forager[t] // group_size
            for (newx, newy) in G.nodes():
                G.nodes[(newx,

```

```

        newy)]['flower'] += (k_flower - G.nodes[
            (newx, newy)]['flower']) * inc
for h in hives:
    for _ in range(np.int32(d_num)):
        nx.set_edge_attributes(G, 1, 'weight')
        curx, cury = h
        omega = 0
        for _ in range(
            np.int32(c_season[t % 365] * f_reps)):
            # for i in range(1):
            a_tired = v_visit / (v_visit + omega)
            s_deduct = 1 / (1 +
                paths[(curx, cury)][0][h]**2)
            G.nodes[(curx, cury)]['flower'] -= (
                1 - np.random.random() * (1 - r_lower)
            ) * group_size * p_pol * a_tired * s_deduct * G.nodes[
                (curx, cury)]['flower'] / k_flower

            # find next node
            adj_loc = [
                n for n in G.neighbors((curx, cury))
            ]
            min_weight = np.inf
            for next_x, next_y in adj_loc:
                min_weight = min(
                    min_weight,
                    G.edges[(curx, cury),
                        (next_x, next_y)]['weight'])
            next_loc = [
                n for n in adj_loc
                if G.edges[(curx, cury),
                    n]['weight'] == min_weight
            ]
            next_x, next_y = next_loc[np.random.randint(
                len(next_loc))]
            omega += G.edges[(curx, cury),
                (next_x, next_y)]['weight']
            G.edges[(curx, cury),
                (next_x, next_y)]['weight'] += 1
            curx, cury = next_x, next_y
res = nx.get_node_attributes(G, 'flower').values()
if (sum(res) < 0.01 * k_flower * 20 and
    max(res) < 0.05 * k_flower):
    print(hive_count, sum(res), max(res))
    return hive_count
else:
    print(hive_count, sum(res), max(res))
    hive_count += 1

```

References

- [1] Jarriault, D., & Mercer, A. R. (2012). Queen mandibular pheromone: questions that remain to be resolved. *Apidologie*, 43(3), 292–307.
- [2] Carrington, D. (2013, February 28). Loss of wild pollinators serious threat to crop yields, study finds. *The Guardian*; *The Guardian*.
- [3] Medicine, C. (2019). Helping Agriculture's helpful honey bees. Retrieved November 15, 2022, from <http://www.fda.gov/animal-veterinary/animal-health-literacy/helping-agricultures-helpful-honey-bees>
- [4] Randall, B. (2022, June 06). The value of birds and bees. Retrieved November 15, 2022, from <http://www.farmers.gov/blog/value-birds-and-bees>
- [5] The need for bees. (2020). Clemson University. Retrieved November 15, 2022, from <http://www.clemson.edu/extension/pollinators/apiculture/importance.html>
- [6] Nearly one in 10 wild bee species face extinction in Europe while the status of more than half remains unknown. (2020, February 20). IUCN report. Retrieved November 15, 2022, from <http://www.iucn.org/content/nearly-one-10-wild-bee-species-face-extinction-europe-while-status-more-half-remains-unknown-iucn-report>
- [7] Sánchez-Bayo, F., & Wyckhuys, K. A. (2019). Worldwide decline of the entomofauna: A review of its drivers. *Biological Conservation*, 232, 8-27.
- [8] Burkle, L. A., Marlin, J. C., & Knight, T. M. (2013). Plant-pollinator interactions over 120 years: Loss of species, co-occurrence, and function. *Science*, 339(6127), 1611-1615.
- [9] The colony and its organization. Mid-Atlantic Apiculture Research and Extension Consortium. (2021). Retrieved November 6, 2022, from <https://canr.udel.edu/maarec/honey-bee-biology/the-colony-and-its-organization/>
- [10] Huang Z-Y, Robinson GE (1996) Regulation of honey bee division of labor by colony age demography. *Behavioral Ecology and Sociobiology* 39: 147–158.
- [11] Cramp D (2008) *A Practical Manual of Beekeeping*. London: How To Books. 304 p.
- [12] Harbo, J. R. (1993). Effect of brood rearing on honey consumption and the survival of worker Honey Bees. *Journal of Apicultural Research*, 32(1), 11–17.
- [13] Khoury, D. S., Barron, A. B., & Myerscough, M. R. (2013). Modelling food and population dynamics in honey bee colonies. *PLoS ONE*, 8(5).
- [14] Russell, S., Barron, A. B., & Harris, D. (2013). Dynamic modelling of Honey Bee (*apis mellifera*) colony growth and failure. *Ecological Modelling*, 265, 158–169.
- [15] Chen, Y. P., Pettis, J. S., Corona, M., Chen, W. P., Li, C. J., Spivak, M., ... Evans, J. D. (2014). Israeli acute paralysis virus: Epidemiology, pathogenesis and implications for honey bee health. *PLoS Pathogens*, 10(7).

-
- [16] University of Georgia. (2018). Managing bees for pollination. Managing Bees for Pollination - Protecting Pollinators. Retrieved November 12, 2022, from <https://bees.caes.uga.edu/bees-beekeeping-pollination/pollination/pollination-managing-bees-for-pollination.html>
- [17] Plantura. (2022). Bee pollination: How does it work? Plantura Magazine. Retrieved November 12, 2022, from <https://plantura.garden/uk/insects/bees/bee-pollination>
- [18] Missouri Botanical Garden. (2013). Bloom Times by Month. Retrieved November 12, 2022, from [https://www.missouribotanicalgarden.org/Portals/0/Gardening/Gardening Help/PDFs/Bloom summary by month.pdf](https://www.missouribotanicalgarden.org/Portals/0/Gardening/Gardening%20Help/PDFs/Bloom%20summary%20by%20month.pdf)