

# Analysis of Factors Affecting the Honeybee Colony Size from the Model-based Perspective

Yibo Ma <sup>1</sup>, Yilei Wang <sup>2</sup>, Evan Zhang <sup>3</sup>, Yuhin Jeremy Tu <sup>4</sup>

<sup>1</sup> Ningbo foreign language school, Ningbo, 315048, China

<sup>2</sup> Zhenjiang Maple Leaf bilingual School, Zhenjiang 212000, China

<sup>3</sup> Vanke Meisha Academy Shenzhen 518000, China

<sup>4</sup> Wuhan Yangtze International School, Wuhan 430101, China

**Abstract.** Bees and other important animals are vital to human survival on the earth. They not only produce delicious honey, but also play an indispensable role in plant pollination, making a unique contribution to the world. The bee population reduction crisis is known as a global problem of sustainable development goals. To consider the factors that can influence the size of honey bee populations and how these factors can affect the size of honey bee populations, we will use the most basic evaluation model. Given the need to accurately calculate the weight of each factor, we use the Analytic Hierarchy Process (AHP) model. In order to analyze the results of the AHP model, we will list our variable factors, apply the AHP model and use sensitivity analysis on it. Also, to predict how many honeybees hives you will need to support pollination of an 81,000 square meters parcel of land containing crops that benefit from pollination. we consider to use Markov Chain Monte Carlo (MCMC) model, which can indicate a stationary distribution to the attendance rate of the bees after n times.

**Keywords:** Honeybee; Gray Forecast Model; Analytic Hierarchy Process; Markov Chain Monte Carlo.

## 1. Introduction

Honeybees, a distinct species known for its social pattern, honey and their venomous stings. They make distinctive contribution to the world not only by producing tasty honey but playing an indispensable role on plants pollination. However, without bee colonies the world is leaded into chaos because they are the most efficient pollinators that way better than human. Honeybees' colony are found facing major challenges from humans' activities, climate changes and invasion of alien species more or less in wild and beehive all over the world [1-3]. The importance of maintaining a stabled honeybee population is important under those external factors due to the significant role it played in this society [4-6]. The honeybee population decline crisis are known as a global issue of SDG and gained eyesight from the United Nations to world governments after numbers of further research conducted. In central America, scholar Marcos have waged several projects in protecting the native stingless honeybee from alien bee(wasps) importation and by protecting the local honeybee they revived the local crops return from the depression started from previous years [7-10]. Yet, the problem isn't solved, the beehives' design and the size of it have few more steps form perfect. To fully tickle the problem, to analyze the negative factors of honeybee colony and the significance of them are important because then we can find ways to encounter them.

Our team consider the factors changing honeybee population in multiple insights, we examine the change of population over time to first confirm the change of natural. Considering that the prediction of population is a grey system, we collected the bee population in the United States from 1969 to 2017 (every four or five years) and improved the grey prediction model. We consider that life span, egg laying rate, fertilized egg/unfertilized egg ratio and other factors will affect the size of bee population, so we apply AHP model and carry out sensitivity analysis. Still, we simulate and predict how many honeycombs are needed to support pollination on a 20 acre (81000 square meters) land. It is necessary to forecast the bee attendance rate and abstract the model considering the land area and bee workload.

## 2. Assumptions and Justifications

1. Given the diversity of bee species, only widely distributed populations of bees with no particular conditions of existence will be considered in this question. As in the case of the bees in the USA in question.

2. In considering the multiple influences, we will assume that the population size of bees is influenced only by the factors mentioned below, Viruses, Pesticides, Predators, Habitat, Lifespan, and Egg laying rate.

3. The occurrence of contingencies will be ignored.

4. Suppose 81000 square meters of land is a square with a side length of about 285 meters, and the activity range of bees is a circle with a radius greater than or equal to 6 km. Bees can pollinate successfully without accidents.

## 3. Honeybee Colony Size Over Time

### 3.1 GFM Model

Indeed, research into the prediction of honeybee populations is important for the conservation of the species. To predict honeybee populations, in this case, we will use the GFM model, based on the honeybee population in the United States from 1969 to 2017 (every four or five-year intervals). GFM model has a high accuracy in predicting short-term future data based on a small number of samples.

### 3.2 Method implementation

Set the known reference data to be:

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)) \quad (1)$$

Do 1 time accumulation (AGO) to produce to form a series:

$$\begin{aligned} x^{(1)} &= (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)) \\ &= (x^{(1)}(1), x^{(1)}(1), x^{(2)}(2), \dots, x^{(1)}(n-1) + x^{(0)}(n)) \\ x^{(1)}(k) &= \sum_{i=1}^k x^{(0)}(i) \quad (k=1, 2, \dots, n) \end{aligned} \quad (2)$$

Find the mean value series:

$$z^{(1)}(k) = 0.5 x^{(1)}(k) + 0.5 x^{(1)}(k-1), \quad k=2, 3, \dots, n \quad (3)$$

Thus, we can get:

$$z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)) \quad (4)$$

And then we construct the differential equations

$$x^{(0)}(k) + a z^{(1)}(k) = b, \quad k=1, 2, \dots, n \quad (5)$$

We construct the GFM model:

$$\hat{x}^{(1)}(k+1) = (x^{(1)}(1) - b/a)e^{-ak} + b/a, \quad k=1, 2, \dots, n-1 \quad (6)$$

### 3.3 Results and Analysis

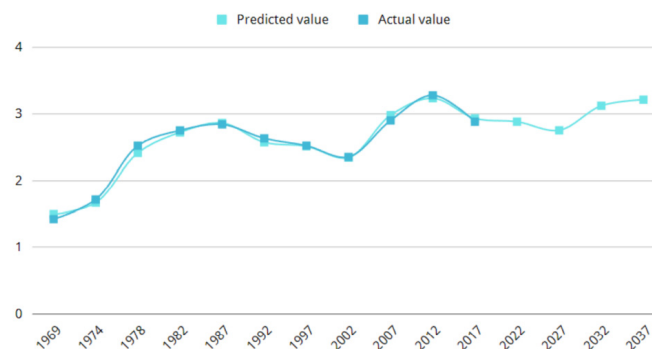


**Figure 1.** The prediction of honeybee populations(million)

Our model predicts that the annual honeybee population in 2022, 2027, 2032 and 2037 will be 2.881, 2.751, 3.121, 3.212 million respectively. As the figure 1.

For prediction class models, the test of data accuracy is essential. For our model an accuracy less than 0.35 is acceptable for the results. As the table below.

**Table 1.** The accuracy of result

| Predicted value | Actual value | Relative Error $e(k)$ |
|-----------------|--------------|-----------------------|
| 1.491           | 1.42         | 0.252                 |
| 1.662           | 1.71         | 0.117                 |
| 2.412           | 2.52         | 0.237                 |
| 2.720           | 2.75         | 0.282                 |
| 2.862           | 2.84         | 0.212                 |
| 2.571           | 2.63         | 0.243                 |
| 2.512           | 2.52         | 0.218                 |

## 4. The Factor of Honeybee Colony Size

In order to consider the factors that can influence the size of honey bee populations and how these factors can affect the size of honey bee populations, we will use the most basic evaluation model. Our model will also include viruses, pesticides, predators, habitat, lifespan, and egg laying rate as factors that can affect the size of honey bee populations. Given the need to accurately calculate the weight of each factor, we finally decided to use the Analytic Hierarchy Process (AHP) model. In order to analyze the results of the AHP model, we will list our variable factors, apply the AHP model and use sensitivity analysis on it.

### 4.1 Deterministic Variables

Certain variables in colony size predetermine whether the bees can live. These are the basic requirements for the honeybee colony size, and they must be met as prerequisites. Because we are primarily concerned the organism, one important consideration is the basis for living. If a group of honeybees cannot be able to live, they cannot have colony size. Another possible deterministic variable is habitat prerequisite. However, our team decides to put that as a probabilistic variable, since in our opinion, in different places, honeybees will have unlike colony sizes, which are based on the factors listed below.

### 4.2 Probabilistic Variables

In this section, we consider variables that do not predetermine the conditions that bees cannot live. They are among the most important variables for the size of population of bees based on sources of bees recently.

**Viruses:** Viruses can affect the colony size of honey bees, as they can cause a rapid decline in the number of individual bees in an area, and in extreme cases, the entire honey bee population can be wiped out. At the same time, bee viruses are not as common in nature [11-13].

**Pesticides:** Pesticide as a by-product of human development also affects the honey bee population. The colony size of honey bees is likely to be limited by the spread of pesticides [14]. Bees may also die due to pesticide contamination. However, pesticides are not common in areas far from cities.

**Predators:** Predators that target honeybees, such as Meropidae, Indicatoridae and Tyrannidae, are also the considerable factor [15-17]. The presence of these predators will undoubtedly limit the development of the colony size of honey bees. Also, these predators are only active in natural areas away from cities.

**Habitat:** Habitat is an extremely important factor in the colony size of honey bees. It plays a decisive role in both urban and natural areas. With today's urban development, the amount of honey

bee habitat around urban areas has decreased dramatically. At the same time, natural honey bee habitats also have an impact on honey bee colony size.

**Lifespan:** Lifespan is an important factor in determining the colony size of bees in areas far from cities. In contrast, this factor was less influential in areas with high urbanization, but was still related to colony size.

**Egg laying rate:** Egg laying rate determines the rate of replenishment of individual bees and is a variable factor to be considered. However, the effect of egg laying rate on honey bee colony size is relatively small compared to other factors.

### 4.3 AHP Model

Given the different performance of each factor in the two cases of proximity and distance from the city, we decided to discuss the weights of each factor in two cases. And, due to the difference between these two cases, the obtained results may be much different. The results obtained are still calculated by the AHP model with the sensitivity analysis. Our hierarchy structure is summarized below.

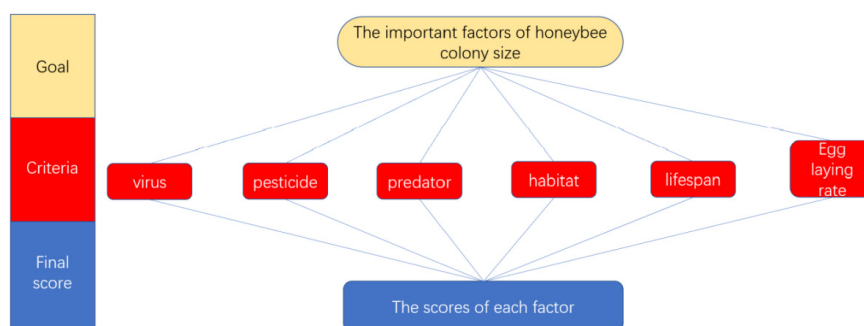


**Figure 2.** Structure of AHP model

Situation 1: The honeybees live far away from the city.

**Table 2.** Factors and Weights

| Factors         | weights  |
|-----------------|----------|
| Viruses         | 0.086205 |
| Pesticides      | 0.055259 |
| Predators       | 0.320405 |
| Habitat         | 0.183709 |
| Lifespan        | 0.289237 |
| Egg laying rate | 0.065185 |

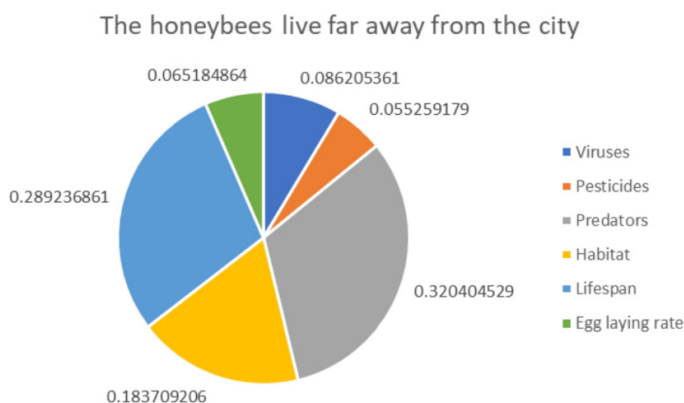


**Figure 3.** The honeybees live far from the city

Situation 2: The honeybees live far next to the urban.

**Table 3.** Factors and Weights

| Factors         | weights  |
|-----------------|----------|
| Viruses         | 0.180422 |
| Pesticides      | 0.198311 |
| Predators       | 0.055837 |
| Habitat         | 0.408311 |
| Lifespan        | 0.100217 |
| Egg laying rate | 0.056902 |

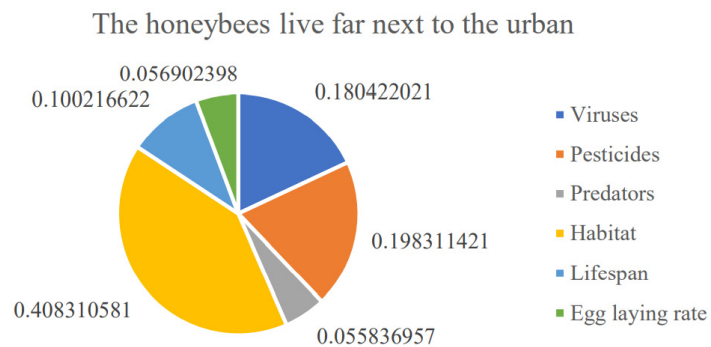


**Figure 4.** The honeybees live far next the urban

#### 4.4 Results and Analysis

In remote areas with low urbanization, predators and lifespans are the most important factors. At the same time, habitat still matters. In areas close to the city, habitat is most important, viruses and pesticides are secondary, and others have little or no importance. For example, because of human activities and development, habitat reduction is serious, pesticides and viruses spread. To combine the results above, we can gain the final weights of each factor, as the table below:

**Table 4.** Factors and Weights

| Factors         | weights  |
|-----------------|----------|
| Viruses         | 0.133314 |
| Pesticides      | 0.126785 |
| Predators       | 0.188121 |
| Habitat         | 0.296001 |
| Lifespan        | 0.194727 |
| Egg laying rate | 0.061044 |

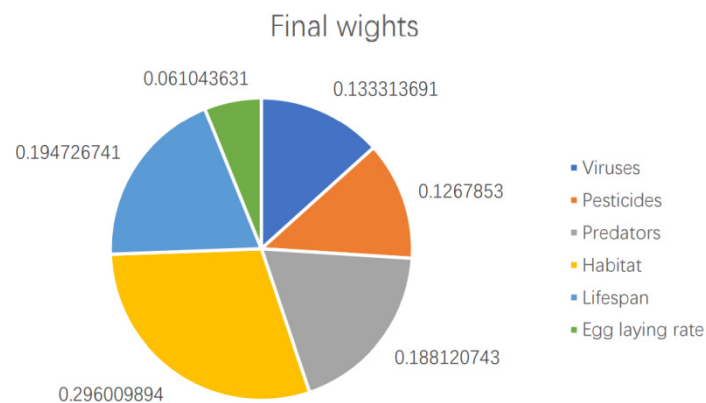


**Figure 5.** Final weights

Thus, the most important factor is the habitat. Lifespan and predator are the next most important influencing factors.

#### 4.5 Results and Analysis

AHP model builds a hierarchy to cover the variables. It applied by portion the weights on the different factors, which compares the importance of plenty of factors. As a result, the importance of each factor is determined by scoring and the model showing the relative importance of each factor to each other.

**Table 5.** Scale and Definition

| Scale      | Definition                                          |
|------------|-----------------------------------------------------|
| 1          | factor i is equally important as factor j           |
| 3          | factor i is slightly more important than factor j   |
| 5          | factor i is apparently more important than factor j |
| 7          | factor i is strongly more important than factor j   |
| 9          | factor i is extremely more important than factor j  |
| 2, 4, 6, 8 | the intermediate values                             |
| Reciprocal | importance scale between j factor and factor i      |

After we ensure the relations between the scales and the definitions, it is simple to build a comparison matrix, which can be expressed as the following from. For each individual, the importance of  $Q_i$  relative to  $Q_j$  is indicated as  $Q_{ij}$ , due to the above scale table.

Furthermore, in order to obtain the value of CI of an  $n \times n$  matrix with the greatest eigenvalue  $\lambda_{\max}$ , we use the expression as the following:

$$CI = (\lambda_{\max} - n) / (n - 1) \quad (7)$$

After getting the values of CI, we are going to find the criteria of the RI. Finally, we can check if CR is always less than 0.1.

$$CR = CI / RI < 0.1 \quad (8)$$

**Table 6.** Average random consistency index

| n  | 3    | 4   | 5    | 6    | 7    | 8    |
|----|------|-----|------|------|------|------|
| RI | 0.58 | 0.9 | 1.12 | 1.24 | 1.32 | 1.41 |

To test our model, we use MATLAB software to practice the data from two situations. In consequence, we confirm that CR is always lower than 0.1. Therefore, our model is acceptable and accurate.

## 5. MCMC Prediction of Beehives

### 5.1 Conditions

Condition 1: We decided to set flowers at a density of 35 flowers per square meter through relevant knowledge. This means that there will be 2,835,000 flowers planted in 81,000 square meters of land.

Condition 2: We know that there are about 2000 bees in a beehive. Not all of these bees can be pollinated, so we finally decided to assume that the number of bees in a hive that can pollinate is 500.

Condition 3: Since the number of flowers pollinated is uncertain during the day, we decided to assume that each flower needs to be pollinated 4 times a day. The pollination of bees does not overlap, which means that ideally a bee that can pollinate 2000 times a day can pollinate exactly 500 flowers.

Condition 4: For the sake of calculation, let's assume that 81,000 square meters of land is a square with a side length of about 285 meters. The range of bees is a circle with a radius greater than or equal to 6 km, so we consider to set the hives outside the square with a suitable distance as the figure 5.1:

Condition 5: The initial attendance rate of the bees is 0.7.

### 5.2 Data Reduction

**Table 7.** Data and quantity

| Definition                          | Quantity         | Symbol   |
|-------------------------------------|------------------|----------|
| Area of Square                      | 81000 mm         | S        |
| Density of Crop                     | 35/mm            | D        |
| Total Crops                         | $S \cdot D$      | T        |
| Pollinated Bees                     | 500              | B        |
| Pollinated Rate                     | 2000/day         | $R_p$    |
| Needed Pollinated Times             | 4/day            | N        |
| Total Needed Pollinated Times       | $N \cdot T$      | P        |
| Initial Attendance Rate of the Bees | 0.7              | $R_{Ai}$ |
| Attendance of Bees after n times    | Unknow           | $R_{An}$ |
| Total Pollinated Bee                | $R_{An} \cdot B$ | $T_p$    |
| Total Pollinated Rate               | $T_p \cdot R_p$  | $T_R$    |

### 5.3 MCMC Model

To define  $R_{An}$ , we consider to use Markov Chain Monte Carlo (MCMC) model. Which can indicate a stationary distribution to the attendance rate of the bees after n times. (1) Determine the prior distribution of the parameter values.

(1) Determine the prior distribution of the parameter values.

(2) Determine the first value of the first visited (or sampled) parameter as the current parameter value.

(3) Based on the value of the currently accessed parameter, propose the value of the next parameter to be considered for access in a certain way.

(4) Compare the posterior probability of the current parameter value and the value of the parameter to be considered for access in the form of a ratio. The calculation of the posterior probability involves the probability density of the prior and posterior probabilities. Depending on the magnitude of this

ratio, the parameter value to be considered for sampling is accepted or rejected. After acceptance, the parameter value is considered as the current parameter value.

(5) Repeat (3) and (4) until some termination condition is met (e.g 10,000 parameter values are visited).

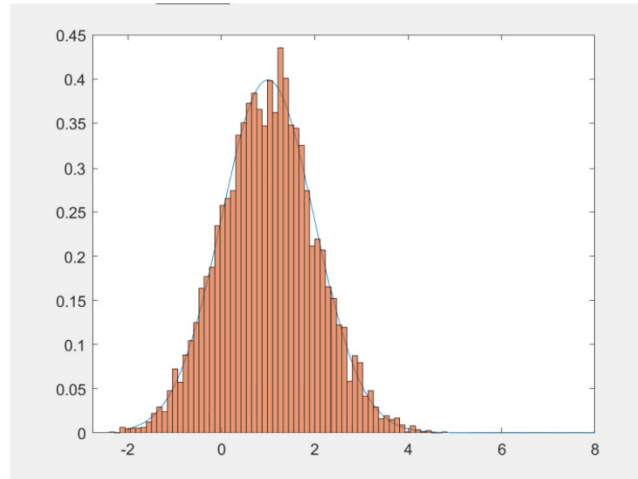


**Figure 6.** the expected value of a complex probability distribution

#### 5.4 Equations of Hives Demand

As we know, the quantity of total crops equals to the area of square times the density of crops, so we get the equation below:

$$T = S \times D \quad (9)$$

As a result, we get the number of total crops is 2,835,000. At mean time, we can use the equation below to identify the final pollinated bee is:

$$T_p = R_{An} \times B \quad (10)$$

In addition, the total pollinated rate is:

$$T_R = T_p \times R_p \quad (11)$$

On top of that, the total needed pollinated times is:

$$P = S \times D \quad (12)$$

Therefore, the demand of beehives is:

$$P / T_R \quad (13)$$

#### 5.5 Consequence

To solve the equation, we finally get the consequence that the final demand of beehives is approximately 28. As the figure below:

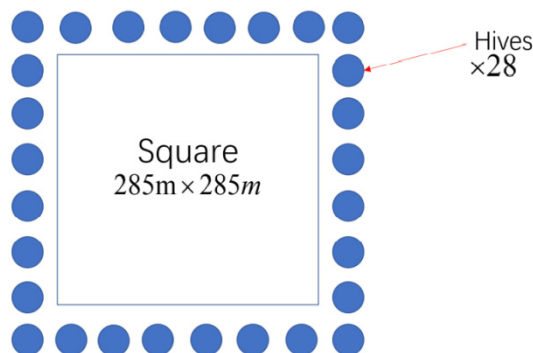


**Figure 7.** The final demand of beehives

## 6. Conclusion

### 6.1 Strength of Our Model

The model takes the research object as a system, and the point of treating the research object as a system is to not cut off the influence of each factor on the result.

This aspect of the model is shown as the weight setting of each layer in the mode ultimately affects the result directly or indirectly, and the degree of influence of each factor in each layer on the result is the same. This makes the method especially useful for systematic evaluations of unstructured properties.

Furthermore, the Process also requires less quantitative data information. Since mode is a method of simulating the way of thinking in a person's decisionmaking process, the mode leaves the step of judging the relative importance of each element to the person deciding the weights, and only keeping the person's impression of the elements, thereby simplifying the process of calculating weights

### 6.2 Weakness of Our Model

The Analytic hierarchy process functions by selecting the best option out of the given alternatives. However, this function only allows the mode to choose from one of the available solutions and cannot provide decision-makers with new solutions to solve problems.

When we apply the mode, our creativity may restrict the quality and quantity of the available alternatives and create a situation where although the optimal solution was chosen, it was still not good enough.

In today's evaluation of scientific methods, it is generally believed that science needs more rigorous mathematical demonstration and quantitative methods. However, the mode requires more thought from human brains than mathematical demonstrations. Thus, the weights decided by one decision-maker may not be as convincing for another person.

## References

- [1] Usda, A. "Insights into social insects from the genome of the honeybee *Apis mellifera* Erratum: 2006 Nov. 23, v. 444, no. 7118, p. 512.]. " Nature (2006).
- [2] Claudianos, C., et al. "A deficit of detoxification enzymes: pesticide sensitivity and environmental response in the honeybee." *Insect Molecular Biology* 15.5(2010):615-636.
- [3] Beye, M., et al. "The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. " *Cell* 114.4(2003):419-429.
- [4] Hammer, M., and R. Menzel. "Learning and memory in honeybee." *The Journal of Neuroscience* 15.3 Pt 1(1995):1617-1630.
- [5] Estoup, A., M. Solignac, and J. M. Cornuet. "Precise Assessment of the Number of Patriline and of Genetic Relatedness in Honeybee Colonies." *Proceedings of the Royal Society B Biological Sciences* 258.1351(1994):1-7.
- [6] Arnaud, et al. "Precise Assessment of the Number of Patriline and of Genetic Relatedness in Honeybee Colonies." *Proceedings of the Royal Society B Biological Sciences* (1994).
- [7] Viktoria, et al. "Primary structures of the N-linked carbohydrate chains from honeybee venom phospholipase A2." *European Journal of Biochemistry* (1993).
- [8] Nazzi, and Francesco. "Synergistic Parasite-Pathogen Interactions Mediated by Host Immunity Can Drive the Collapse of Honeybee Colonies. " *PLoS Pathogens* (2012).
- [9] Higes, M., et al. "How natural infection by *Nosema ceranae* causes honeybee colony collapse. " *Environmental Microbiology* 10.10(2010):2659-2669.
- [10] Martinson, V. G., J. Moy, and N. A. Moran. "Establishment of Characteristic Gut Bacteria during Development of the Honeybee Worker." *Applied & Environmental Microbiology* 78.8(2012):2830-2840.

- [11] Farooqui, T., et al. "Modulation of early olfactory processing by an octopaminergic reinforcement pathway in the honeybee. " *Journal of Neuroscience the Official Journal of the Society for Neuroscience* 23.12(2003):5370-80.
- [12] Randolph, et al. "Cognitive architecture of a mini-brain: the honeybee -ScienceDirect." *Trends in Cognitive Sciences* 5.2(2001):62-71.
- [13] Menzel, R. G. "Searching for the memory trace in a mini-brain, the honeybee. " *Learning & Memory* 8.2(2001):53.
- [14] Tessier, D. C., et al. "Enhanced secretion from insect cells of a foreign protein fused to the honeybee melittin signal peptide." *Gene* 98.2(1991):177.
- [15] Durst, C., S Eichmüller, and R. Menzel. "Development and experience lead to increased volume of subcompartments of the honeybee mushroom body. " *Behave Neural Biol* 62.3 (1994): 259-263.
- [16] Tarpy, D. R. "Genetic diversity within honeybee colonies prevents severe infections and promotes colony growth." *Proceedings of the Royal Society B: Biological Sciences* (2003).
- [17] Kuchler, K., et al. "Analysis of the cDNA for phospholipase A2 from honeybee venom glands: The deduced amino acid sequence reveals homology to the corresponding vertebrate enzymes." *European Journal of Biochemistry* (1989).