Apicultural Conservation and Technology: An Investigation into Decreasing Bee Populations with Data Science

Christopher Petrucelli

Michael Gildein, Faculty Advisor

Spring 2021

Table of Contents

Part 1: Apis Mellifera

Introduction

Since the Centers for Disease Control and Prevention's (CDC's) initial publication of disappearing bee colonies in 2006, public attention towards the dangers faced by pollinators like honeybees has grown tremendously both on traditional and social media platforms. This trend has resulted in countless calls to action, requesting that people support local apiaries and keep pollinators in mind when applying pesticides or herbicides. They are not wrong about the danger, North America has 4 species of bumblebee in decline, and one that has gone extinct. Throughout all this media craze, though, it would seem that the general population learned little about the pollinators that they advocated for or how they could be helped. There are several interesting projects that want to do more, such as the Pollinators Pavilion, an electronically monitored habitat for wild pollinator species that is being deployed locally in the New York Hudson Valley, but this type of project is a bit out of reach of a student thesis project. With my background in computer science, I looked towards projects that were more theoretical, and could be done without actually keeping any bees, and after enough bumbling, landed on beehive monitoring technology.

This project serves two purposes: the first is to increase understanding of pollinators, specifically the Apis mellifera (or European/Western Honey Bee), the most commonly kept type of honeybee in North America. Throughout this paper, the lives and functions of honeybees and their beehive, as well some of the common threats that they face will be covered. This is to provide the reader a basic understanding of how pollinators live and of their importance in the environment. The second part of the project, motivated by the research in the first, will apply my Computer Science background to beekeeping in an investigation into how technology is, and could be used to protect beehives. It will explore commercially available beehive monitoring technology, existing beehive data metrics, and new ideas to discover what data is, and can be gathered from apiaries, and whether there is good reason to spend the money to do so. The completed analysis of beehive metrics and their uses will support the discussion of a theoretical 'best' beehive monitoring solution, that incorporates all the metrics determined to be useful to making sure that the kept bees are healthy and productive at a reasonable cost.

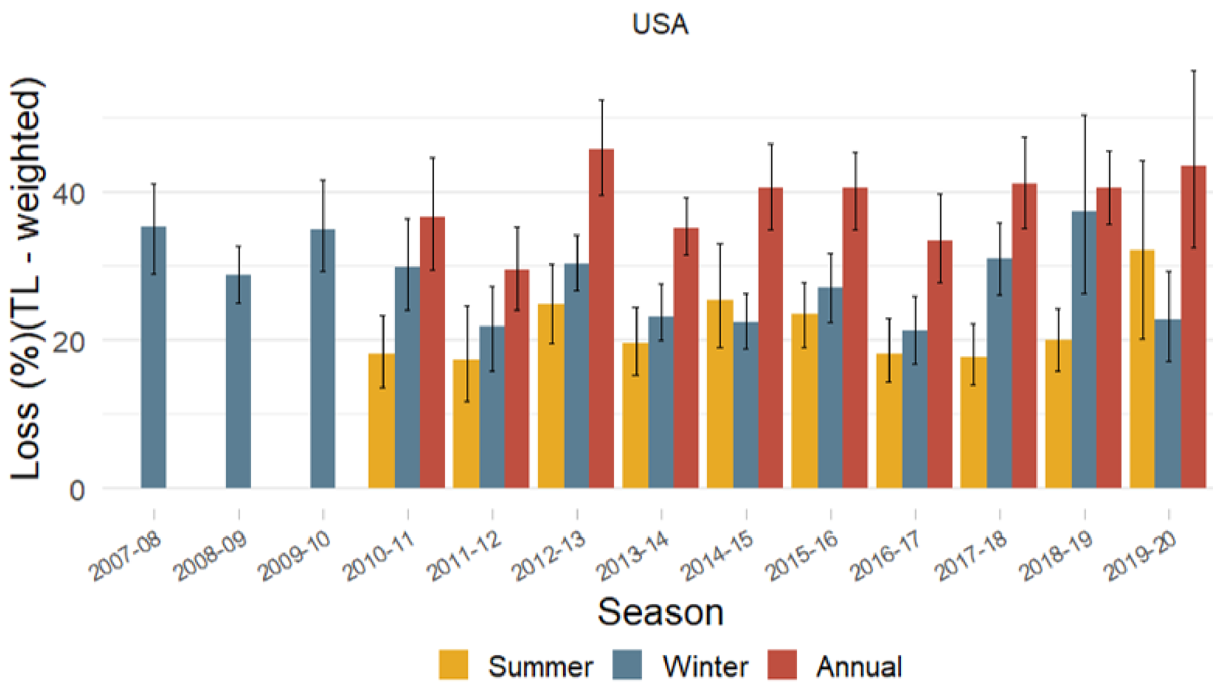

Fig 01: Graph of beehive losses in the U.S. since 2006. The data here began in 2006 as a survey created by the Apiary Inspectors of America (AIA), and soon became the basis for a U.S. Department of Agriculture grant under the umbrella of the Bee Informed Partnership.

Background

The beekeeping profession has a very diverse nomenclature, so it is important to define some terminology that will be used. A group of bees being kept by a beekeeper can be referred to by a few names, but this paper will use ‘bees’ or ‘colony’. This is to create a distinction between the bees themselves and their housing, which will be referred to as a ‘beehive’, ‘hive’ or ‘box’. This is a generalization, as there are many varieties of hives used in beekeeping, but it is not valuable to distinguish them in this paper. In a typical hive of any variety, the main body pieces of the hive are known as ‘supers’. There are usually a number of supers in a hive, with the largest sitting at the bottom, making up the core of the hive, and a handful of smaller supers on top, separated from the large bottom super by a wire frame called the ‘queen excluder’. The queen excluder simply prevents the queen from moving up to those supers. A super contains a number of ‘frames’, which are spaces upon which the colony will construct wax cells. These cells can fill a variety of purposes, such as food storage or brood development. The ‘brood’ is the young of the colony, which will be explained in greater detail during discussion of the bee life cycle. The brood frames are all located in the largest (usually bottom) super. Figure 02 shows the location of these pieces of a hive in a Langstroth-style hive.

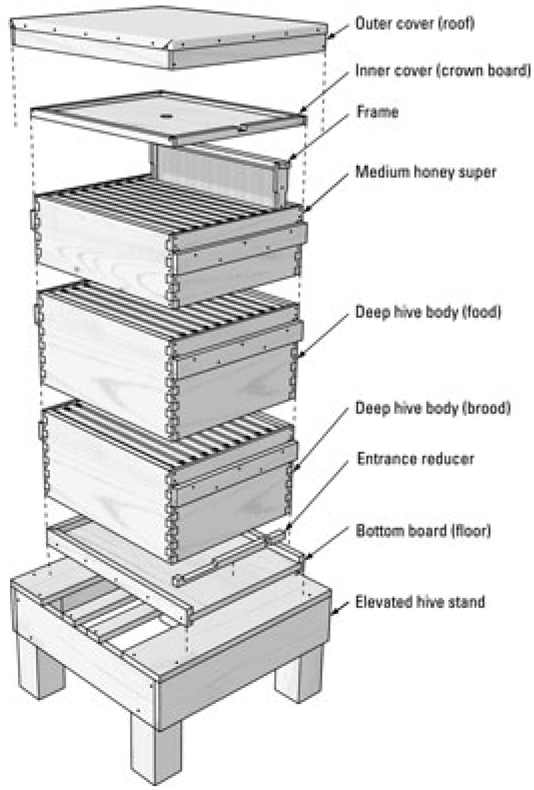

Fig 02: Diagram of a Langstroth style hive. The queen excluder would be positioned above the deep hive body and below the honey supers.

Bee Basics

Apis mellifera is the most commonly kept type of bee in North America, though the species originated from Europe/the Middle East/Africa and was spread due to the economic benefits of pollination and honey. In fact, this species of bee is on every continent but Antarctica, although different species of bees are kept in other parts of the world (Such as Apis cerana in South-East and mainland Asia). Bees have haplo-diploid sex determination at birth, meaning that an egg that is not fertilized will become a male 'drone' bee, and one that is fertilized will become a female 'worker' bee. The worker bee, while female, cannot reproduce, as the single 'queen' bee is the only egg-laying bee in a colony. Each of the different types of bee in the colony has different roles to perform. The worker bees take on the most diverse workload, they are responsible for tending to the brood, building wax comb, handling food storage, guarding the hive, and of course, foraging outside the hive for pollen and nectar. Drone bees, on the opposite end of the spectrum, only have one real task, which is to mate with a queen from another hive, and do not actually have a stinger. Finally, the queen bee spends her early life mating, and uses what she gathered in those days to lay up to 1500 eggs a day for the hive, for the rest of her life. All this activity is coordinated through chemical signals called pheromones, and all the types of bees produce some type of pheromone. Some of the interesting ones are the queen mandibular pheromone, which informs the colony of the queen's location, the brood pheromones that signal what type of care to provide to growing bees, and the Nasanov pheromone, which coordinates the location of the colony to workers that got lost. Everyone's favorite thing about bees, their stings, also release a pheromone that attracts more bees to come and sting the same target, but stings are used in a purely defensive manner, except for the queen.

Fig 03: Drone bee (left), Worker bee (right)

Brood Development

Eggs are laid by the queen into hexagonal wax cells in an area of the hive called the ‘brood area’, or in apiaries, often on a 'brood frame'. The distinction between an egg hatching a drone or a worker bee was explained, but the determining factor between getting a queen and worker bee lies in what diet the larvae receive: workers receive a standard diet of some royal jelly (a substance produced by worker bees that nurse the larvae), and then pollen/nectar, while queen bee larvae eat only royal jelly. All bees go through a metamorphosis from egg, to larvae (also called 'open brood' because the wax cells are uncapped), to pupae (also called 'closed brood' because the wax cells are capped), and finally to an adult, but do it in different amounts of time. Queens are the fastest at around 16 days, next is workers at around 21 days, and finally drones at about 24 days to mature.

Fig 04: Adult (left), Pupae (center), Larvae (right)

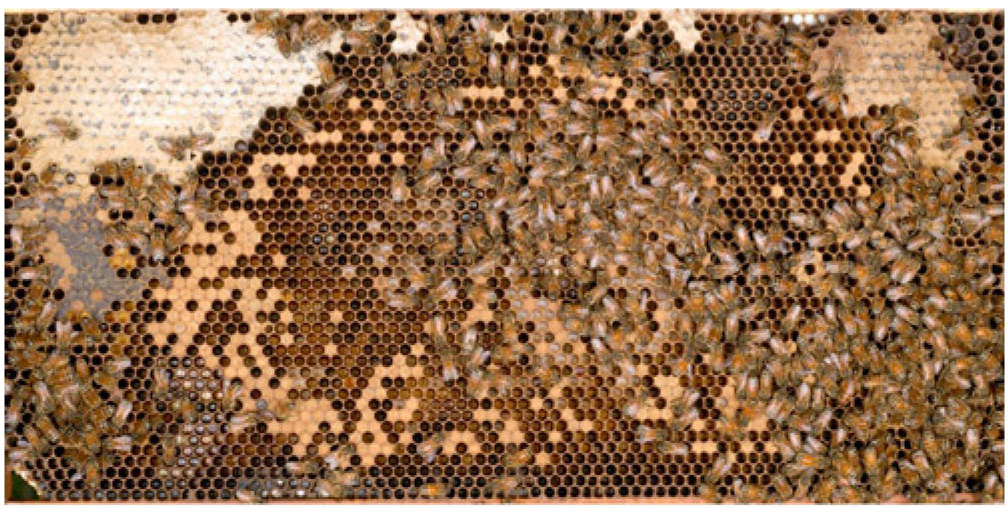

Fig 05: Brood frame with bees in all stages of development

Lifecycle

Now that we know a bit more about the life of bees inside the hive, let's talk about the general life cycle that bees will take year-to-year. Every Spring, the time when food sources are most abundant, a colony that survived Winter will attempt to create a new colony (in a new hive) in a process called 'swarming'. The colony first produces 10-20 new queen candidates, and when they are close to being fully developed, the previous queen will take around 2/3 of the worker bees from the hive and leave to find a place for a new colony. This is a difficult trip, and it is very possible that it will fail, but the initial colony will be sustained by one of the new queens until she can repeat the process again next year. Speaking of the new queen, there are two ways that the new queen is determined from the 10-20 candidates. The first way is that one of the queens emerges from their closed brood cap earlier than the others, and kills them before they can emerge as well. Otherwise, when they emerge closer to the same time, the queen candidates will fight until there is only one survivor, who becomes the queen.

Once the new queen is established, things continue much as they always did in the old hive: the queen mates and lays eggs, while worker bees work on all of the various things that they do in the hive. The Western honeybee is suited to temperate climates, and only during the spring is food plentiful enough to undertake swarming. The rest of the year is spent stockpiling honey, produced when nectar from flowers goes through enzymatic processes inside bees, and is dehydrated and stored in capped wax cells. This honey will be the colony's food source during the cold/wet times in the year, and is ultimately the only source of food to get the colony through winter. One exception to this is if the colony is in an apiary and can be fed supplements by the beekeeper, but this is only done when absolutely necessary.

Threats

There are several threats to the Western honeybee, and while any one of them might be a death sentence for wild bees, kept bees have humans on their side. Common threats include disease, parasites, pesticides, and temperature and environment, but others, like long-distance transportation, are less obvious threats to bee health.

Most diseases that can disastrously affect the Western honeybee often target the brood, but there are some diseases that can affect fully developed bees as well. Some illnesses, like Chalkbrood and Sacbrood, are easily handled and can often be disposed of by the worker bees themselves. Worse is European Foulbrood, which kills the brood in their larval stage, but this disease can be cured with medical treatment and cleaning of the brood frame. The deadliest disease for a hive is American Foulbrood, which works in much the same way, but is resistant to the only approved drugs to combat it. If it gets too far in a hive, the colony is doomed. All of the brood illnesses can be detected early by physically examining the brood cells' color and texture, and by also examining bees in all of the stages of development. For adult bees, the most common illnesses are all treatable and preventable. Nosema and Amoeba interfere with the bees' ability to digest pollen, dramatically shortening lifespan and food production, but can be treated with antibiotics. To fully prevent spread from recurring, a beekeeper would need to replace the combs on the frames of their beehives. The other common illness is Dysentery, which could be caused by a number of things, but ultimately has similar effects on bees as it would on humans. While Dysentery is easily identified by brown/yellow streaks on the top bars and inner cover of the hive box, it's even easier to avoid it altogether by maintaining hygiene in the apiary.

For parasites, there are a few that could impact a colony, but I want to specifically mention two, Varroa mites and Wax moths. Varroa mites are tiny pests that can coexist with Eastern honeybees, but not with Western honeybees. The mites do not do much themselves, but their bites can pierce the skin of the Western honeybee, allowing diseases to spread much more quickly than they otherwise might have. Among other treatments to prevent Varroa mite infestations, a hive usually has a screen filter along the bottom, which allows the tiny mites to fall through when they die, while the bees themselves are unable to fit. It is important, especially in the late Spring and Summer, for beekeepers to check on the tray under this screen to monitor mite numbers. Wax moths take a more direct approach to damaging a colony by eating brood or wax, but are similarly treatable when impacting a hive. They are detected by the silky webs left behind in the wax comb that they damage.

Another threat, which has been in the spotlight since 2006, is known as Colony Collapse Disorder (CCD). Around that time, beekeepers began reporting massive losses of between 30-90% of their colonies, all of which showed some unexplained new phenomena. In CCD cases, the population of worker bees in the affected colonies suddenly drops to almost none, with few dead bees being present in or near the hive. Even more confusing is that the queen and brood remain in the hive, and in many cases honey and pollen reserves were plentiful. Unfortunately, without worker bees, these affected colonies have no chance of surviving. There are a number of possible explanations put forward, including disease, pest infection, pesticide use, changes to the environment, and others, but the true cause is still unknown, and in the case of many of those afflictions, it would be common to see thousands of dead bees near or within the hive. Thankfully, the rise in reported cases of CCD quickly subsided. In 2008, 60% of winter losses were attributed to CCD, but by 2013, that figure had dropped to around 30%. Interestingly, there are agricultural records from over 100 years ago that noted years of colony decline, but it’s impossible to say whether the causes then were the same as CCD.

Importance of Apiculture

Apiculture provides a variety of products and services which impact both the economy and the environment. In the U.S., the value of honey produced over the course of a year has remained at around 300 to 340 million dollars since 2013. While this is the most commonly thought of product, there are some other interesting products and services bees provide as well. There is another substance that bees produce (aside from honey and royal jelly) known as propolis. Propolis, also known as ‘bee glue’, is similar to a resin and is typically used by bees in maintaining, repairing, and weather-proofing the comb of the hive. Humans, however, have used propolis extensively in medicine throughout history and to this day as alternative medicine, as it possesses anti-microbial properties useful in treating wounds, burns, sore throat, respiratory infections, and more. Another way that bees impact the economy is through paid pollination services. These services take hives of bees on the road, and leave them to provide pollination for the customer’s crops. It is estimated that bees pollinate crops worth $15 billion each year in the U.S. alone. This method of pollination is used extensively for crops like blueberries and cherries, but one crop in particular, almonds, are almost 100% reliant on honey bees for pollination.

Part 2: Beehive Monitoring

Introduction

Beginning the exploration into data science, this project has used the Cross Industry Standard Process for Data Mining (CRISP-DM) approach to data mining, preparation, and analysis. At this point, we have described the first step of the six presented in CRISP-DM, business understanding. In this case, business understanding means having an understanding of bees, what they do, threats they face, etc. This type of information is important because it can help provide context and meaning to the actual data gathered from hives. The remainder of the paper will focus primarily on the second and third steps of CRISP-DM, data understanding and data preparation, as well as some of the fourth and fifth steps, modeling and evaluating the data.

To better manage their colonies, and ensure that they remain healthy and safe, a beekeeper may employ the use of one of any number of beehive monitoring technologies. While all unique in their own right, the similarities and differences between these systems can help to outline what beehive data is valuable for data analysis. The data visualized in this project comes from one of two sources. The first set of data was acquired from Kaggle and spans 3 hives in an unknown location in the US in 2017. This data set contained data on temperature, humidity, weight, and bee flow in and out of the hive. The second set of data was pulled from the BeeCounted Org. interactive map from 8 hives in the Poughkeepsie-Newburgh area. In this data set, there were temperature readings for all of the hives, and humidity and weight readings for only a few of them. The majority of the hives also only had data in limited windows of time, but one hive (id = ‘58ho’) had temperature and humidity data from 2017 to the present. For links to the datasets used, see the datasets section of the paper here.

One setback encountered during this project was the lack of availability of beehive data sources. Obviously, actually having a base of data is a key component of performing data analysis. Only the 2017 data set was found available for download, the 2016-2021 data was pulled from BeeCounted by issuing specific requests for the hive IDs listed in the Poughkeepsie, NY area. Without a large base of data, the ability to use machine learning to train a model, likely towards the end goal of predicting hive health and activity, is going to be severely limited. Other features that others are attempting or that this project will theorize, such as photographic detection of Varroa mite infestations or brood diseases, would require a large amount of image training data in order to be worth the resources to implement. If the solution still requires a beekeeper to visually inspect the hive, what is the return on the cost of implementing cameras into their beehive monitoring solution when they could just check the hive physically? This lack of availability occurs even though there are a variety of existing beehive monitoring solutions. With few exceptions, these companies gather a base of data from their customer’s hives, but do not openly share any of this data publicly. While this project was able to work with what was available, it was also primarily theoretical in nature. Without any data, it may be close to impossible for disruptive products to succeed in the market against solutions which have already assembled a base of data upon which they can train their machine learning models.

It would also be useful to be able to access a base of data from within the same geographic region. It was possible to do that to some extent in this project, using the BeeCounted data, but that data was not made easily accessible, nor was it particularly plentiful. The benefit of having a set of data from hives in the same area is that each hive could potentially be a reference point for another. Hives within the same area (therefore facing similar conditions) and of a relatively similar size would probably exhibit similar activity and production, and potentially face the same dangers. Comparing the activities of a number of hives in this situation would make discerning underperforming hives much easier, as well as ease the need for a baseline of data, like seasonal weight patterns, for new hives.

Data Preparation

Data preparation, the third step in the CRISP-DM cycle, will be discussed first, as the charts used as support for the data understanding and modeling steps were processed and modified beforehand. The first step of the cleansing process was to combine the data from each hive into their own dataframes. During this process, a few changes to the original data occurred. In the BeeCounted data set, the timestamp index was broken down into a *year* and *dayofyear* column, so that data could be grouped and visualized by year (for the few hives that had years worth of data), as shown in figure 06.

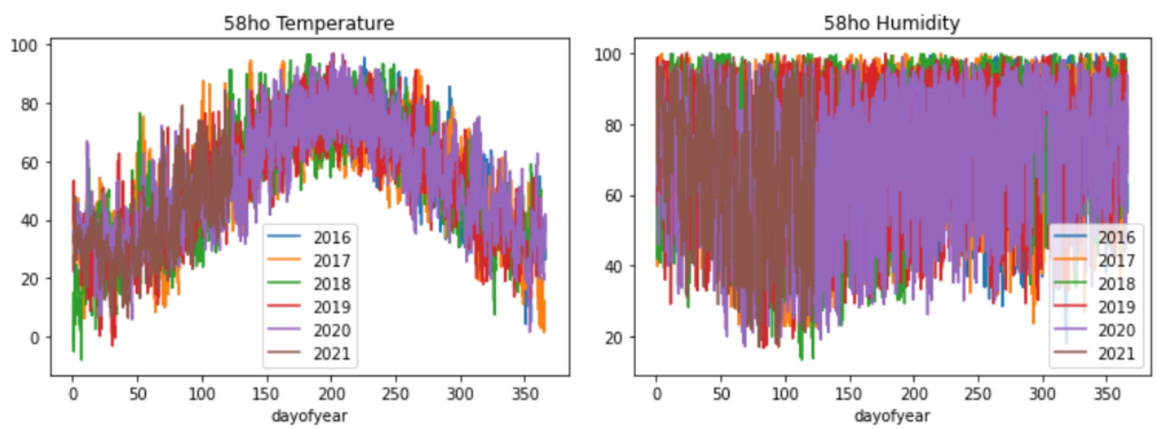

Fig 06: BeeCounted hive 58ho temperature and humidity by year

In the 2017 data set, there were sensor readings from ten different temperature sensors, but unfortunately no indication of what reading came from where. To accommodate this, the temperature values for each timestamp were condensed into a *min*, *mean*, and *max* value at that timestamp, as shown in figure 07.

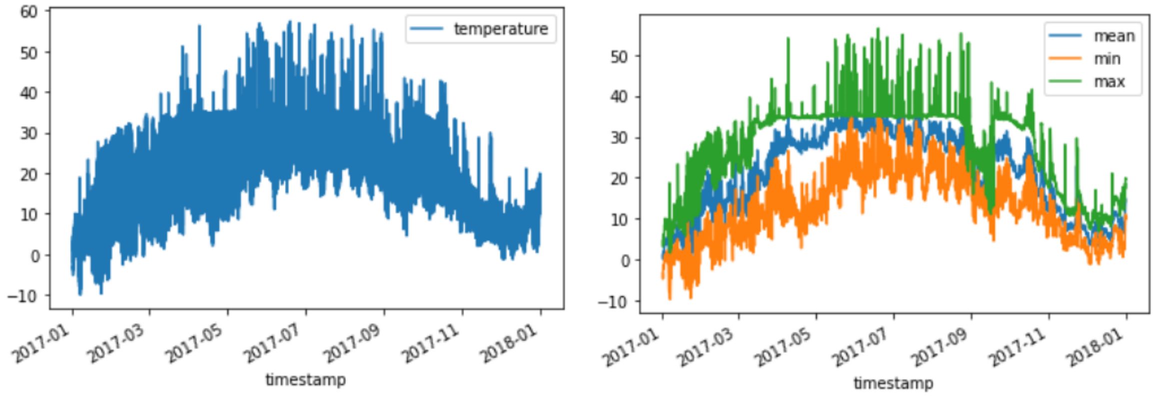

Fig 07: 2017 original temperature data (left) and after modification (right)

Additionally, the flow readings in and out originally existed as a single column of data that had the same timestamp. In this data, a negative value (or 0) indicated the out-flow, and a positive value (or 0) indicated in-flow. To more easily congregate the data together for comparison, these values were merged as *in* and *out*. To avoid having to make these changes every time, the modified data frames were exported to CSV files and replaced the original temperature and flow imports.

Finally, a new column, *weightchange*, was added to track the change in weight from reading to reading. This was accomplished by applying a lambda function to each row of the dataframe, which calculated the difference in weight from the current reading to the previous reading, and assigned that value to the current row's *weightchange* value. This was done primarily because the raw weight data, without contextual information like how heavy the hive itself is, holds very little meaning. Figure 08 shows the *weight* and *weightchange* values measured on May 1st, 2017.

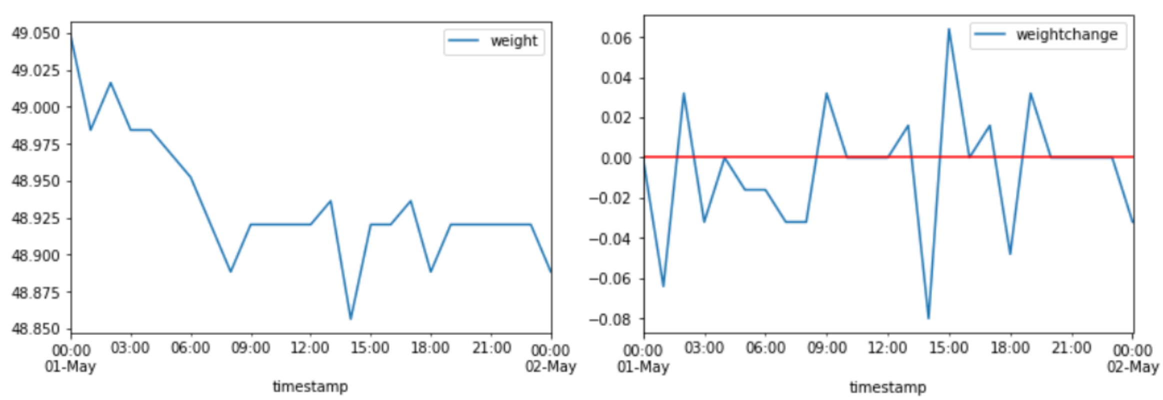

Fig 08: May 1st, 2017 - weight (left) and weight change (right)

The next step in cleansing the data was to remove outlier values. Since there was very little available data, it wasn't in our best interest to remove large amounts of data from either of the data sets, so a good strategy was needed. Values were considered to be outliers if they sat more than 1.5 interquartile ranges (IQRs) outside of quartile 1 or 3. For example, the 2017 data set's mean temperature data had a Q1 = 9.9 and Q3 = 25.3 degrees Celsius. With those numbers, our outlier range would be calculated as:

$$IQR = 25.3 - 9.9 = 15.4$$

$$[Bottom\ range,\ Top\ range] = [9.9 - (1.5*IQR),\ 25.3 + (1.5*IQR)] = [-13.2,\ 48.4]$$

When dealing with something like temperature, weight, and humidity, though, it was often easy to determine outliers simply by the possibility of such a thing occurring in nature. It required no calculation to determine that a temperature reading of -227 degrees celsius, or a humidity reading of -66% were very likely to be sensor errors. Alternatively, we could accomplish this by comparing external and internal temperatures, as it is not possible for the internal temperature to be colder than the external temperature. Using the IQR method of detecting outliers did cause some issues, though. Specifically regarding humidity data, the range of values (1.5*IQR) outside of quartile 1 or 3 would sometimes include valid values that were more extremely low or high. The solution to this problem was to not use the IQR method, but rather just ensure that humidity values were between 0 and 100. The same problem occurred with temperature data, so temperature values between -10 and 110 Fahrenheit (or -23 and 44 Celsius) were considered valid. Another option for detecting outliers is to use Z-scores. A Z-score is the number of standard deviations outside the mean of the data a single data point is, typically between -2 and 2 or -3 and 3. These values are used because on a normally distributed curve, 2 to 3 standard deviations away from the mean will encapsulate more than 95% of the values.

The final step taken in cleaning the data was to determine how best to deal with missing data. With the limited quantity and timeframe of data available, it is unfortunate that there were a number of blank spots where the sensor did not report any readings, but that is the nature of inexpensive and compact sensor systems. In the 2017 data set, there were sensors that took readings more frequently, or otherwise on different schedules than others, and thus introduced a few nulls when merged. In the BeeCounted data set, for many hives the range of timestamps that had values for one sensor was often not the same as the range of timestamps for other sensors, likely because different sensors were being added into the hive monitoring system at different times. Had the data been consistently reported over the course of a few years, the best solution would likely be to generate estimated values based on the trends of the past years. Finding a good variable for regression would be a common way to do this, as you could then use the model to predict missing values based on the value of the dependent variable. In consideration of this, a linear regression was performed using each column of data as the independent variable, and every other column as the dependent variable. With the available data, this did not produce any meaningful R^2 scores (Specifically, high R^2 scores), but the limited timeframe of the data may be preventing the model from recognizing seasonal trends.

Data Understanding, Modeling, and Analysis

Now that the data has been cleaned, it is time to take a closer look at step two of the CRISP-DM model, data understanding, to understand what data is gathered, why, and what conclusions can be made using the data gathered. The most common feature of the beehive monitoring solutions explored was internal temperature and humidity measurements. Bees are able to carefully control the temperature of their hive, which is especially important during periods of brood development, by utilizing either friction to heat up (by vibrating their bodies) or a sort of air conditioning to cool down (by flapping their wings). A temperature reading from inside the box during brood development should be within the 90-95 degrees Fahrenheit, or 32-35 degrees Celsius range. Values outside this range, especially if taken close to the brood frame, may indicate a non-laying queen, queen-lessness, lack of resources, or preparations for swarming. Regarding external temperatures, bees are very good at handling a wide range of external temperatures so long as they are not wet, but there are merits to recording external temperatures. Foraging activity will be lowered the colder/wetter it is outside, and a long string of days like this might impact the hive's ability to gather sufficient resources, requiring supplemental feeding by the beekeeper.

Knowing that worker bees closely monitor the temperature of the hive for brood development, it would follow logically that the colony's ability to gather resources is more limited when external conditions require them to heat/cool the hive. However, that might also imply that activity would be increased when conditions inside the hive are within an acceptable range. Using the 2017 hive data set, figure 09 shows the comparison of the mean temperature across all sensors to the number of bees in/out to test this hypothesis.

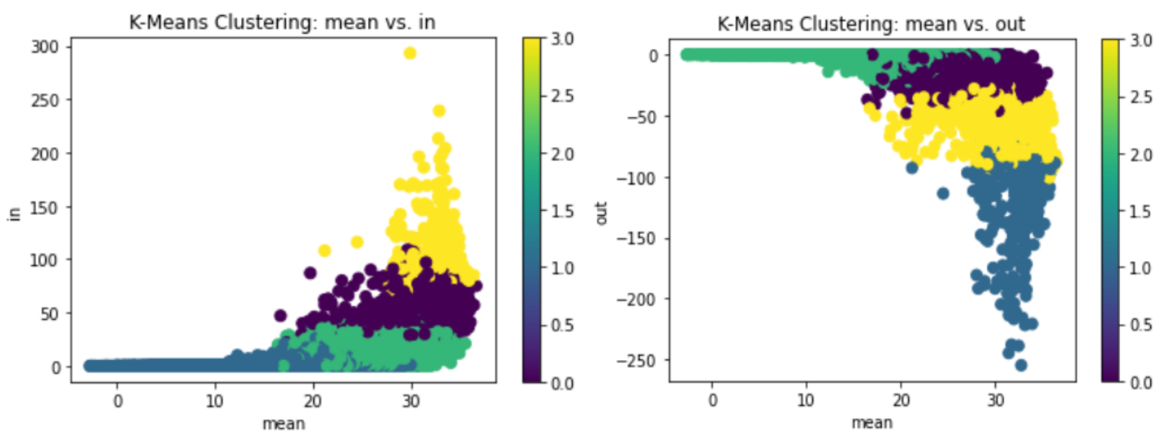

Fig 09: K-Means Cluster using a K value of 4, plotting mean temperature against hive flow. A value of K = 4 clusters was used because it converged into distinctly grouped bands.

As the hypothesis predicted, the peak for bees in/out occurs when the mean temperature is in the 30-33 degree Celsius range. Although there isn't much data for mean temperatures further above 33 degrees Celsius, there is still enough data available to see the spike level-off as the temperature gets to 33 degrees and beyond. These graphs also show that bees start leaving the hive to gather once the temperature rises above about 15 degrees Celsius. Although the data set did not specify the location of sensors, we see an even clearer spike in figure 10 when using the max temperature recorded by a sensor rather than the mean.

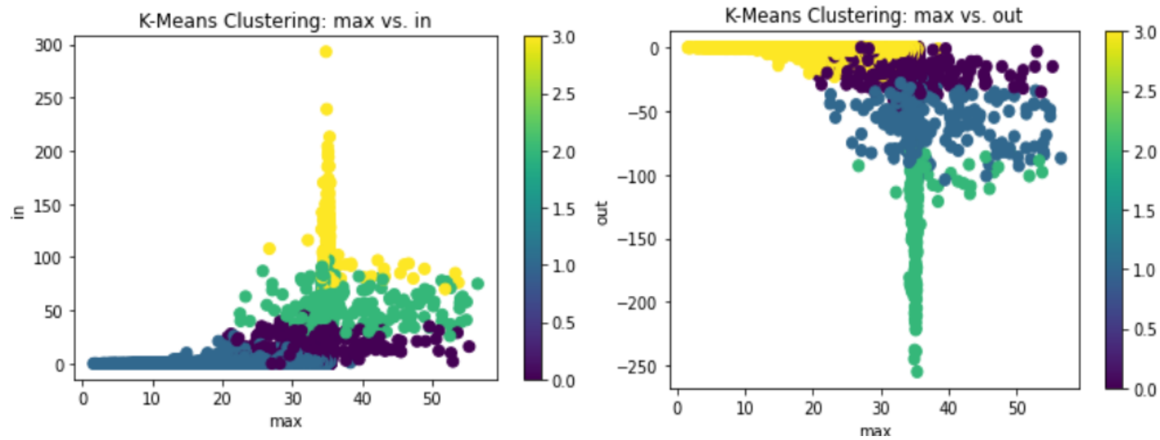

Fig 10: K-Means Cluster using a K value of 4, plotting max temperature recorded on a sensor against hive flow

This hypothesis being proven correct might have some interesting applications for further experiments, provided one had access to a colony. One such experiment would be to attempt to aid the bees in regulating the temperature of the hive, perhaps with something simple like providing the hive shade when the temperature is greater than 30 degrees Celsius and in direct sunlight. If it is possible, using methods like this one and others, to keep the brood frame in the correct temperature range with less intervention from the worker bees, then it may be possible to see gathering activity more consistently at the values seen in the spike. Without actually performing the experiment, though, we can't know what effect disturbing the hive will have, so perhaps the proving of this hypothesis might serve as a warning to not disturb the hive when it is in the optimal temperature range so that the bees are able to freely maximize their gathering activity.

Another common metric that is captured is humidity, again both internally and externally to the hive. Similarly to temperature, humidity is managed primarily for the brood. During brood-less periods, the internal humidity is allowed to fluctuate with the temperature, but during brooding, humidity is optimally kept between 50-60%, but can comfortably sit between 40-80%. An interesting reason to capture this metric, along with external humidity, is to compare how much changes to the external humidity impact internal humidity. If the internal humidity fluctuates significantly when external humidity changes, that may indicate that the hive is not healthy. The same reasoning can be used when comparing internal and external temperatures. A high internal humidity can also reduce Varroa mite reproduction rates, as well as increase occurrences of chalkbrood. If a hive is suffering from either of these afflictions, you might be able to predict improvement or worsening of the colony's condition, but other measures should be taken as well.

The 2017 data can be used again to see if a similar relationship exists between humidity and in/out as it did for temperature and in/out. Assuming this is the case, then we can hypothesize that bees in/out should be at its peak within the 50-60% humidity range, but otherwise high between the 40-80% range.

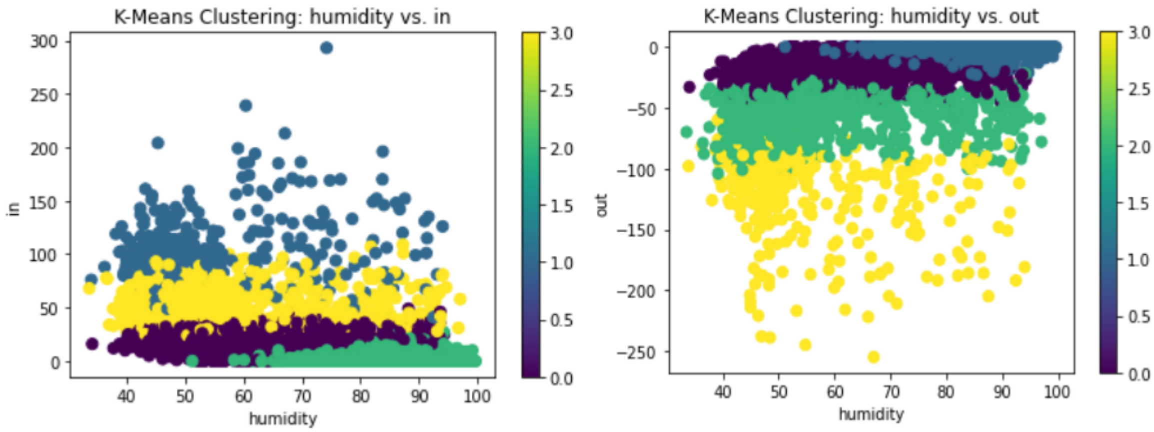

Fig 11: K-Means Cluster using a K value of 4, plotting humidity against hive flow

From these charts, it doesn't seem like the relationship between humidity's optimal range and in/out flow is anywhere near as clear as between the optimal temperature range and flow. This does make sense, though, as the restriction on keeping the brood temperature within a range of 3 degrees Celsius is much more strict than keeping the humidity within the 40-80% range. Beyond that, there is a limited amount of data outside the 40-80% range in question. It seems like activity may slow down at close to 100% humidity, but there is not enough data below 40% to see if the same may be true on the lower end of the range.

Measuring the weight of the hive is something that not many of the examined beehive monitoring solutions included, but it is still a useful metric. Weight measurements will allow the beekeeper to determine roughly how much food a hive has accumulated during their active season, as well as how much the hive has consumed during the winter months. While this metric is strongly dependent on each beekeeper's circumstances (hive type, size, wax comb development) and the conditions of the year, generally a hive in a warm area that experiences a short winter should have a minimum of 40 pounds of honey stored for winter consumption, and a hive in a colder area with a regular winter season should have a minimum of 30 to 60 pounds to as much as 80 to 90 pounds stored. During the active season, a hive should still have a minimum of 10 pounds of stored honey at any given time. Monitoring food stores is not the only benefit to measuring hive weight, though. Another reason to measure weight is that sudden increases in hive weight may be indicative that a different hive has collapsed, and the bees from this hive are looting the other. This is a dangerous situation, as an infestation or disease that wiped out that other hive may be spread to this one by the worker bees who visit it to loot. A consideration to make, though, is that not all sudden weight increases necessarily mean that the bees are looting a collapsed hive. For example, a wooden hive will have a sudden increase in weight after it rains, as the wood absorbs water, and then a subsequent drop as the box dries. Although unavoidable to some extent, choosing the correct materials for your hive can help to avoid this problem. Oils in the wood of cypress and cedar trees make it naturally resistant to weather and bug infestation over more common (and inexpensive) types of wood, like pine. Another popular option is to use synthetic wood, created from recycled plastics. This material is completely weatherproof and will not rot, making it a fantastic choice for a new hive.

Other Sensors and Data

There are a few other sensors worth mentioning that were utilized by some, but not all of the beehive monitoring solutions. Light sensors within the lid of the hive, accelerometers, and E-compasses were used by some solutions as a means of theft or damage prevention. Simply sending alerts for changes to any of these sensor's readings would serve to inform the beekeeper that a person or animal might be interfering with the hive. The next metric gathered by some solutions was barometric pressure. Like other animals, bees can detect changes in barometric pressure better than humans can, and you can expect less activity at times of low pressure, such as when a storm is approaching. The main reason to measure barometric pressure is that some beekeepers have reported that their colonies are far more agitated than normal during periods of low pressure. This did not seem to be the case in every situation, but it might be better to stay away from your hives until the pressure normalizes, unless you are looking for a few stings. Cameras or other types of motion sensors can be used to monitor bee activity outside the hive, at least as far as monitoring bees the number of bees entering and exiting the hive. This metric is similar to weight in that it will be more useful for a hive that already has a baseline to measure against. However, monitoring bee movement is another way to potentially be warned of any issues with the colony or a need for supplemental feeding, as low activity may lead to a lack of resources within the hive. Finally, some monitoring solutions have begun to capture audio readings from inside the hive, and have used AI to analyze this audio. According to these solutions, audio readings may be able to predict when the queen has stopped laying eggs, when a new queen has been born, or that swarming is soon to begin.

Over the course of this exploration of beehive monitoring systems, it has become apparent that there are a number of metrics that an experienced beekeeper would know to gather from their hives manually. Metrics such as brood area temperature and weight can be found in hive record cards, which are logs made to record the status of a hive during periodic checkups by the beekeeper. Others, like knowing that hives may become irritated during periods of low barometric pressure, could also be learned over time, or through a hard lesson or two. For the experienced beekeeper, it may not be worthwhile to invest in a beehive monitoring solution that does not include more than those basic metrics that they are already familiar with. For an inexperienced beekeeper, though, the basics offered by a beehive monitoring solution are sure to be helpful not only in actually monitoring their hives, but also in understanding the information so that they are less reliant on the technology.

While compiling information on beehive monitoring technologies, a few new ideas for using sensors or data came up. One involves Varroa mites, and the aforementioned screen and space that sits at the bottom of the hive to collect the dead mites and monitor the level of infestation. By taking a picture of the collection area daily, the task of monitoring for these parasites can be integrated into the hive monitoring solution. The size of the space, as well as cost considerations, will limit the quality of images that can be taken, which may be problematic when attempting to visually distinguish the tiny mites in the collection area, but something as simple as changing the color of the bottom of the area to blue will help the small orange mites stand out against the background. Similarly, a camera could be used to monitor the brood frame for abnormalities that might indicate a disease is spreading in, or that a parasite is destroying the brood frame.

Another application for the E-compass, which has otherwise been an anti-theft device in hive monitoring solutions, is to prevent a phenomenon known as 'drifting'. Drifting is an issue that is unique to bees in an apiary, especially in commercial settings when there are many hives in one area. When a worker bee first begins to go out to forage, she leaves the hive and flies in a spiral pattern, getting further and further from the hive, but always looking inwards towards the hive entrance. The purpose of this is to become familiar with the location of the hive, but when multiple hives near to one another all face the same direction, the bee may accidentally begin returning to the wrong hive. If a number of hives are all in a line, again facing the same direction, the hives on the ends will tend to receive a large number of drifting bees, and therefore have significantly more honey than the other hives. Placing hives in a circle, limiting the number of hives on a line to two, or placing hives more randomly and facing in different directions will help to prevent this phenomenon. Using the E-compass readings from hive monitors in a large group of hives, the facing of each hive can be compared, and an alert could be raised if the hive configuration is likely to cause significant drifting.

Finally, although it will not assist in discovering the cause of CCD, there may be a way to detect a case of CCD. As had been mentioned, CCD cases are particularly strange because the population of workers in the affected hives do not die, but suddenly disappear. Whether by photographic or other means, monitoring the bee flow out of the hive could detect the mass migration of workers out of the hive. Perhaps by knowing more about when the workers leave or where they go, more can be discovered about the phenomenon.

Conclusion

Having completed a basic investigation into understanding bees, exploring a handful of beehive monitoring solutions, and gathering, cleansing, and modeling what beehive data was available, the only way to conclude this project is to come back around and assess the effectiveness of beehive monitoring solutions in ensuring the health and safety of our bee populations, as that was one of the original questions posed in the project. From the research and analysis done in this project, beehive monitoring solutions are a great way to keep hives healthy. While it is true that experienced keepers could likely manage without a beehive monitoring system, amateurs are sure to benefit from the extra layer of oversight on their colonies. Even with an experienced keeper, though, a major advantage of beehive monitoring systems is that they are able to monitor the hives at all hours of the day, and can alert the keeper when anything abnormal occurs. Additionally, with a large enough base of data, machine learning will be extremely valuable in recognizing trends in data that a human would overlook.

Looking to costs, a tech-savvy beekeeper could probably create their own beehive monitoring system that includes basic features (such as temperature, humidity, pressure, light, etc.) for less than \$100 using a small computer like a Raspberry Pi. Of the beehive monitoring solutions examined, the least expensive costs \$160 per hive, and some solutions can range into thousands of dollars per hive before including monthly subscription plans to use their visualization tools and apps. The advantage of using an existing system is that, as said previously, they do have an established base of beehive data to train machine learning models, and many have already done so and have implemented them into the hive monitoring tool or app associated with the solution. Thankfully, there are organizations such as BeeCounted that are trying to amass and make beehive data available to the public, because without buy-in from a number of beekeepers, a new hive monitoring solution is not likely to be able to gather enough data to even keep up with existing products.

Although the open availability of data is currently limited, there is a bright future for continued advancements in hive monitoring technology. Beekeeping is yet another field where technology can be employed to increase productivity, understanding, and potential indicators of opportunities for assistance. Audio analysis of hives is just one example of the field being pushed further, and there will surely be more advances in predicting colony activity. Hopefully the time is not far off that hive monitoring systems will be able to detect pheromones produced by the colony or by the queen, and incorporate those metrics into their analyses. For the time being, though, beehive monitoring solutions have absolutely begun the process of modernizing beekeeping, and that work is important to the protection of a vulnerable and environmentally irreplaceable species.

Works Cited

"Beehive Scales and Monitors." *Beehive Scales and Monitors*, 9 May 2019, https://solutionbee.com .

Blackiston, Howland. "How to Choose Lumber for Your Beehive." *Dummies*, www.dummies.com/home-garden/hobby-farming/beekeeping/how-to-choose-lumber-for-your-beehive/ .

Blackiston, Howland. "The Parts of a Beehive." *Dummies*, www.dummies.com/home-garden/hobby-farming/beekeeping/the-parts-of-a-beehive/.

"BuzzBox Hive Health Monitor & Beekeeping App." *OSBeehives*, www.osbeehives.com/.

"Colony Collapse Disorder." *EPA*, Environmental Protection Agency, 26 Apr. 2018, www.epa.gov/pollinator-protection/colony-collapse-disorder .

"Innovative Beehive Monitoring System by Pollenity." *Pollenity*, 13 Feb. 2020, https://pollenity.com.

Jones, Richard, and Sharon Sweeney-Lynch. *The Beekeeper's Bible: Bees, Honey, Recipes & Other Home Uses*. Stewart, Tabori & Chang, 2011.

Kočalka, Peter. "How To Use A Beehive Entrance Reducer?" *Bee Hive Monitoring*, www.beehivemonitoring.com/en/blog/post/how-to-use-a-beehive-entrance-reducer.html.

"Loss & Management Survey." *Bee Informed Partnership*, https://beeinformed.org/citizen-science/loss-and-management-survey/ .

Mortensen, Ashley N, et al. *European Honey Bee - Apis Mellifera*, University of Florida, Aug. 2013, entnemdept.ufl.edu/creatures/misc/BEES/euro_honey_bee.htm.

"Remote Beehive Monitoring." *Arnia*, www.arnia.co/ .

Shahbandeh, M. "Value of Honey Production in the U.S. 2020." *Statista*, 29 Apr. 2021, www.statista.com/statistics/191993/value-of-honey-production-in-the-us/#:~:text=According%20to%20the%20report%2C%20the,million%20U.S.%20dollars%20in%202020.

"Threats to Bees." *Museum of the Earth*, www.museumoftheearth.org/bees/threats.

Wagh, Vijay D. "Propolis: a Wonder Bees Product and Its Pharmacological Potentials." *Advances in Pharmacological Sciences*, Hindawi Publishing Corporation, 9 Dec. 2013, www.ncbi.nlm.nih.gov/pmc/articles/PMC3872021/ .

"Weighing Your Beehive: Why and How." *PerfectBee*, www.perfectbee.com/blog/weighing-your-beehive-why-and-how#:~:text=A%20medium%20super%20will%20weigh,should%20weigh%20about%2080%20lbs .

Datasets

BeeCounted Data Set: https://map.beecounted.org/citizen_science/embedded_map

See the *beecounted.ipynb* file in the project Github repository for CURL HTTP requests made to gather the data, as well as cleansing steps and additional modeling.

Beehive Metrics 2017 Data Set: https://www.kaggle.com/se18m502/bee-hive-metrics

See the *beehive-metrics_2017.ipynb* file in the project Github repository for data cleansing steps and additional modeling.

Github Repository

https://github.com/AFineSortie/Apicultural-Conservation-and-Technology