Data Analysis for Beehive Monitoring System

by

Nadia Khoshgoo

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

Master of Science

Department of Electrical and Computer Engineering

University of Manitoba

Winnipeg, MB

Acknowledgements

Abstract

We can increase the production of honey from honeybees, if we optimize the conditions affecting its production in their hive. To do this, and to enable better decision-making, we need to understand the relationships that affect honey production. This research focuses on understanding the relationship between the various factors that affect honeybee activity and honey production. Towards this goal, various data analysis software programs were developed to find the relationships that exist between the various factors affecting bee activity. One of the best ways to find the relationship between these factors would be to determine the rules that exist between them, and the frequency that they occur.

In this thesis, research data were obtained by implementing various sensors in a beehive. Data is analyzed over a sliding window, and the relationship between data factors are obtained. Various times during day or night are analyzed to determine important data, and data relationships. Generally, it was found that comparing more data within a given sliding window sample, and comparing over smaller window sizes, better determines more common rules as well as global rules amongst the sensor data.

Table of Contents

List of Tables

List of Figures

Glossary Listing and Abbreviations

Frequency: number of occurrences

Min: minimum

Max: maximum

MinFrequency: Minimum frequency

Interval day number: Day number starting from the start date entered by the user.

All-Factors-Relationship rule: The decision-making rule to be defined between the data of various sensors. Sensor data relationships are defined by the "sign" of sensor output changes from a prior amount. Sign for positive change is "P", for a constant one is "S", and for a negative one is "N".

Two-Factors-Relationship frequency: The frequencies of similar changes sign and opposite changes sign of two factors when one is the focus.

PN charts: They show Two-Factors-Relationship frequency outputs indicating PN frequencies (positive-negative changes frequencies).

Capacitance boards: These are the sensors boards that were constructed for use in the hive, and upon which the bees build the honeycombs. They were designed by Valerie Beynon [2].

Arduino Board: A microcontroller board with input and output pins, including I2C ports for sending clock signals and sending receiving data [29].

Adafruit data logger shield: This enables the Arduino board to connect to an SD card [31].

Temperature/Humidity sensors: Adafruit Si7021 [30]. In this sensor SCL and SDA pins were used to be connected to the multiplexer.

Multiplexer: Adafruit TCA9548A 1-T0-8 I2C Multiplexer breakout [32]. This chip is used to connect up to eight I2C sensors to one Arduino I2C port.

Capacitance sensor: Adafruit MPR121 12-Key capacitance touch sensor breakout [35], with SCL

and SDA connection. These capacitance sensor chips can receive 12 capacitance inputs.

Light Sensor: Adafruit APDS9960 Breakout is called as "Adafruit APDS9960 Proximity, Light,

RGB and Gesture Sensor" in Adafruit website [34] -The sensor measures the outdoor

light(lux) for this Phase 2 of the project for this thesis. [34]

Chapter 1. Introduction

1.1.   Content statement

Bees are essential to many organisms in this world, as for example, they provide essential pollination service. Since various factors affect the heath of bees, and as a result their honey production, in order to gain insight into these factors, it is valuable to gather data using a hive monitoring system. In other words, understanding the relationship between factors affecting the hive system can be helpful to beekeepers. In this thesis, various sensors were placed in hives to study hive conditions. Hardware was developed to enable the sensors, which included temperature and humidity monitoring, capacitance boards to monitor bee activity and honey production, and light sensors. Two types of software programs were written to read and analyze sensors data. One was for user interaction, where a user defines parameters for the functions done by the software analysis. The other performs various the data analysis to find the relationships that may exist between the data from each sensor in the hive.

1.2.   Focus and importance of the thesis

In this thesis, the focus is on finding the relationship between the sensors' data obtained from the hive system. The largest data set used in this thesis is the data obtained in 2018, employing some of the sensor hardware developed by Valerie Beynon's prior research. The objective of the software written in this thesis is to identify bee activity changes, and their relation to other factors in the hive. Each sensor (such as a temperature sensor or light sensor) is one factor. Considering that the hive environment and the bees' activities are affected by various factors, the various types

of sensors are considered together to help get a broader knowledge of the hive environment. To find out the relationships between various factors, the developed software uses data mining methods, and identified frequently occurring relationships between data. As the various data parameters are happening together and can affect each other, finding the longest item-sets and their frequencies, can show what are the relationships between various factors. In studying the factors, the sequence of data items read at a point of time is not important and is not considered here, and so it is not a sequenced frequent item-set.

In the analysis, one item at a time is focused upon, in order to find out its "rule" relationship between other factors and their data. To identify differences between general rules and outliers, the frequency of each rule relationship is counted. Because a factor can be changed due to another factor's changes, analysis is done to find and extract the longest rule sentence of frequent factors. The factors can be decreasing or increasing. Therefore, they are also considered as factors. Each time a sensor is focused upon, it is identified when it is maximum or minimum over a current sampling window. The changes of other factors are identified as well, in order to generate rule sentences. The number of time this rule is identified, is frequency, is determined and is added to the rule sentences. The comparison of these rules shows the relationship of the various factors on each other. The most frequent rule can be named as the "dominated rule".

The importance of this thesis is that it demonstrates a way to understand how the hive system is affected by the various factors and how the bees work in various situations. The analysis of the factors relationships to each other, can help to find the important relationship forming rules in the hive system. By identifying these relationships, this knowledge can be used to improve the hive situation, and to understand the activities of the bees within at different times or in response to different situations. The goal of the hardware researched in this thesis is to develop technologies

for remote monitoring of hives. It should be noted that when adding sensors to a hive, it can change the environment itself. Therefore, future study should include the possible effects this may have on the behavior of the bees in the hive.

More generally, the data analysis methods investigated in this thesis, demonstrate how this method can be used for other large data sets, in order to find relationships between datasets.

1.3. Problem statements and objectives

Problem (A): User may need specific data including, average daily comparisons, average and maximums and minimums of a specific factor in the hive, in order to see the daily overall trends and changes in bee activities. However, with thousands number of records of data it is not easy to recognize.

Objective (A): Software is developed to determine the user defined calculation that address this issue by finding averages and slopes.

Problem (B): Analyzing one sensor's data by itself cannot help revealing the main points about the hive, since the relationships between the various factors cannot be easily found.

Objective (B): The various factors in the hive may affect each other. Finding the relationship rules between them can be obtained by analyzing data from a comprehensive hive monitoring system. To do this, hardware was developed to integrate and interrogate multiple sensors in the have, to gather needed data for analysis.

Problem (C): Analyzing the trends of data obtained from the hive system can be challenging because of the thousands of records. Additionally, simply finding the global maximum and

minimum for each sensor data in the system is not enough. Considering the large number of records and data, a global comparison of all data can result in not considering the local relationship between various sensor data.

Objective (C): The software for analyzing of the whole data from, consisting of various sensors of a monitoring system, employs a sliding window technique. Local windows based on time are created, and trends and rules in that period are found. In this way, the rules in sliding windows would be found and the frequency of their occurrences in the whole dataset would be added and categorized by its frequency.

Problem (D): Bees are living creatures. They have rules and habits and will act accordingly regarding environmental changes, or day or nighttime changes. Therefore, understanding their habits and rules would be important to improve the situation of the hive.

Objective (D): To determine the habits of the bees in the hive, in the sliding window of study, various times of day and night are investigated, and the data from the sensors are compared to their prior values. The results of these comparisons identify rule relationships. Therefore, common habits of the bees during a time of a day and night in the same conditions would be identified. If the conditions have changes as determined by which sensor reading is focused upon, this would be considered to create new rules. At the end, the rules having the greatest frequency of occurrence are the ones that can show the habits of the bees.

Problem (E): Outliers may lead to a sudden increase or decrease in the value read by a sensor, and they should not mislead the research.

Objective (E): Outliers would be defined as happening only a few times. Since, their frequencies would be lower than other rules, they would be eliminated by their low frequencies.

Problem (F): Maximization or minimization of some elements can affect hive and bee activity. Therefore, their consideration is still important.

Objective (F): The maximum and minimum of a sensor data can lead to important understanding of bee activity changes, or a change in the hive situation. A focused parameter would be researched when it is found to be minimum or maximum in a sliding window.

1.4. Thesis outline

This thesis has 8 chapters. Chapter 2 discusses some literature about previous hive monitoring systems and gives a background in on data analysis concepts and methods. Chapter 3 discusses phase 1 of thesis which is the first effort at implementing two temperature sensors connected by multiplexer to Arduino and its data acquisition software programs. Moreover, this chapter describes two software programs for data acquisition and data analysis for hive data obtained by Dr. Curry's research group [1] and Valerie Beynon [2]. Chapter 4 is the beginning of phase 2 of thesis. It gives a description of the hardware implemented in Valerie Beynon's M.Sc. research, which is based on the multiplexer system of Chapter 3, and visual description of the data file. The data produced by this sensor system is the main input for data analysis in the chapter 6 and chapter 7. Chapter 5 presents pictures of the capacitance sensor boards during the in-hive studies with the phase 2 sensor hardware. Chapter 6 focuses on the finding all factors relationship between sensor readings in the hive and determines the frequent rules that exist between them when one factor is the focus. Chapter 6 also focuses on the finding Two-Factors-Relationships' frequencies when one

factor is the focus. Chapter 7 focuses on the hive observations and some photos showing the capacitance boards' status during visits, along with some experiments. Chapter 8 is the thesis conclusion and discusses future works.

1.5.   Contributions

- Multiplexer hardware/software design for multi-sensor integration to Arduino sensor system.

- Software development to read and visualize data from capacitance sensor boards.

- Software to import data from Arduino multi-sensor system to data analysis software.

- Development of the algorithms and software for the Two-Factors-Relationship and All-Factors-Relationship data analysis.

- Study of which factors result in rules with the highest frequency of occurrence, in order to determine which factors are important, and their correlation to the hive environment and bee activity.

Chapter 2. Literature Review

2.1.   Honeybees and their environment

Honeybees are important creatures that are necessary for nature. The benefits that they provide nature and human beings are numerous, including the production of the product which is honey that is used both as food and a medical substance [3]. They are also helpful to nature by providing pollination of crops, including many fruits and vegetables [4]. Examples of honey's benefits for health including its use as an antibiotic for treating and disinfecting wounds and burns, and for curing coughs [3].

Honeybees work to maintain their environment above 9-10°C, because temperatures below 8°C can cause inactive honeybees to enter a coma and die according to James Simpson [5]. Simpson states that honeybees gather in clusters to maintain warmth, and the size of these clusters expands in winter and contracts in warmer weather. In cold winter temperatures, the edge of the cluster is sharp. It is also stated that the temperature where the brood is located is the warmest, around 34-35 °C.

According to the thesis of Graham Parsons [6], a problem that beekeeping has faced in Europe, Asia, and Canada is Colony Collapse Disorder (CCD). This is the sudden drop in bee population in a hive, which sometimes reaches 30%. In an attempt to manage and study this problem, environmental sensors and monitoring systems have been placed in hives to monitor what is happening in the hive.

2.2. Hive monitoring systems

Various Beehive monitoring systems have been developed. The system "Ariana hive" gathers data on hive temperature, humidity, weight and bee count, and apiary weather condition. It can compare the data from various sensors together, and the customer can share hive data [7]. Another system is the B-ware System, which also gathers hive weight and temperature data every 15 minutes [8]. The BeePi system uses audio, video, and temperature data to study the beehive. Their system uses USB microphones for audio monitoring [9]. BuzzBox [10] monitors the sounds created by bees for monitoring the health of the hive. The Hive-Tech system [11] uses smell sensors, acoustic sensors, and weight sensors for monitoring hive health. APiS Tech [12] monitors the physical status of hive, and it includes alerts bee counting, monitors wasp attack and queen laying, and send alerts to customers about temperature or weight changes. BroodMinder [13] focuses on monitoring weight and inside hive conditions and provides data comparison. It uses Bluetooth technology and circuits to measure temperature and humidity inside the hive, outside temperature, and measures weight. Moreover, it offers storage of data with no cost in Cloud and the data is open for everyone to access.

Several other recent beehive monitoring devices as are described in the following paragraphs. Firstly, Beebot and uHive by Pollenity [14,15] offer a complete hive structure with temperature and humidity sensor, acoustic analysis, and optional solar panel. Beebot is the sensor system itself, and uHive is includes the hive boxes. Oldooz Pooyanfar from SFU [16,17] has designed a bee health smart system using several microscopic sensors, accelerometer, and microphones, along with temperature and humidity measurement in each hive. They analyze hive sounds, and report the anomalies occurring in the hive in real time. SmartHive [18] is an automated beekeeping

system that assists in collecting honey, by removing the bees safely from each comb. This device can help the beekeeper by reducing labor and increasing honey production.

The smart hive monitoring system being investigated in this thesis employs various sensors, which are connected to a micro-controller. To understand the role of each device in a monitoring system, it can be compared to a human body. In a human body we have brain which is the decision-making unit, in computer system this unit is a microprocessor. We have various parts connected to our brain like the short and long memory. These can be compared to the cache and flash memories that are part of a microprocessor, which is at the heart of a micro-controller. The micro-controller includes various ports that can be connected to devices that can act as input or output, similar to the receptors of a human body. An Arduino micro-controller has ports for connecting various sensors and has input and output communication. Considering various possible uses of this board, other accompanying boards are available that can be connected to it for various reasons, such as reading or writing from SD card. These boards are excellent tools to gather various and continuous data from connected sensors. In addition to the hardware and software developed for the hive monitoring system of this thesis, work was done to enable data analysis of the gathered sensor data.

2.3.   Supervised vs Unsupervised Learning

The following sections introduce concepts in machine learning. Descriptions of supervised and unsupervised learning methods are given.

2.3.1.   Supervised Learning

Supervised learning is a method of machine learning to use to train an AI by using data with correct answers to teach the system or program the right answer. In supervised learning there would be prior knowledge, for example previous input and output that can be referred for learning trends and they would be used for predictions [19]. Some examples for supervised learning are in the below two paragraphs.

- **Supervised Learning in visual recognition:** In [20], the AI was trained using 2 million videos with known answers in order to teach it how to detect a pedestrian. After this was done another set of data was used to check the accuracy of detection.

- **Supervised Learning in sorting:** Supervised learning was used to teach a robot to use visual recognition to separate trash into 12 bins [20]. To give feedback on errors, a human expert examines the bins and point out any mistakes in sorting. The classification which is described in section 2.4.2 is a kind of supervised learning.

2.3.2.   Unsupervised Learning

In unsupervised learning there is only input data information without any prior knowledge, and the data may be divided into different parts [19]. Unsupervised learning deals with data without previous training and it uses similarities to group data [21]. Two of the main categories of unsupervised learning are Clustering and Dimensionality reduction. Clustering searches the data to divide the data into different parts based on the structures that might be the same. Dimensional reduction focuses on getting rid of parts of data that might be unnecessary or might have error, and is also done to make the data set simpler. There are several examples of unsupervised learning in [21] that includes the following methods shown below.

- **Means Clustering:** The Means Clustering method divides data into clusters based on similar structures, and elements of data. Usual functions for this method include anomaly detection, pattern recognition, audio segmentation [21].

- **Hidden Markov Model:** This is a model that uses statistical methods to research the data and to divide the data based on its elements' characteristics. It is a version of simple Markov chains that research data based on the mode of the data, and so it helps enable the data to be researched considering other aspects. Examples where it is used include text translation and speech recognition [21].

- **DBSCAN Clustering:** This refers to Density Based Spatial Clustering and involves looking into data to sort them based on their similarities. An example of its usage is in customized advertisement and for customer service [21].

- **PRINCIPAL COMPONENT ANALSYS (PCA) Data Analytics:** This is a dimension reduction method for data visualization, which shows the divided information based on unique aspects of data. For example, this has been employed for data related to a website visiting [21].

- **T-SNE Data Analytics Visualization:** T-distributed Stochastic Embedding is a method to visualize high dimensional data by a low dimensional way and calculates the data's probability based on their common values in both high dimension and low dimensions. It then tries to decrease the differences between the high dimension and low dimension results by using conditional probabilities. An example of its usage is for complicated non-constant cases like for genome visualization and medical tests [21].

- **SINGULAR VALUE DECOMPOSITION (SVD)- Recommender SYSTEM:** This is used as dimension reduction methods and focuses on some features for specific goals.

Examples include giving suggestion to buyers such as which items of clothing suits another piece of clothing [21].

- **Associative Rules:** Predictive Analytics is a method for predictive goals in e-commerce, such as to find relationships of objects in a transaction. For example, what items are popular and how are they popularly bought one after another [21].

2.4. Data mining techniques

The following sections present various data mining techniques, along with a description of terminologies.

2.4.1. Data cleaning and preparation

Data cleansing is the process of preparing data and acquire the needed data from the raw data [22]. Much of the time of a data scientist may be spent on data cleansing, and some data cleansings includes deleting extra spaces, dealing with blank cells, changing of string type numbers to number fields, remove duplicate records, etc. [23]. Some other descriptions of data cleansing include removing unwanted observation such as irrelevant or duplicate ones, and some other methods may be fixing structure errors like typos, inconsistent capitalizations, etc. [24].

Moreover, filtering unwanted outliers for cases where an outlier is illegitimate or is a quantity that doesn't seem to be based on a real data, helps in the accomplishment of the model routines such as Linear Aggression. Handling missing data may be done by deleting the record having missing data, or inputting it based on other records data. Marking missing categorical data as "missing" can help in research and indicate its consideration in future studies, algorithms, etc. [24].

2.4.2. Classification

Classification is used to identify or categorize data based on data types and data attributes [22]. According to kdnuggets [25], classification is a supervised learning method for prediction. It includes labeling data and categorizing it into classes, then it tries to predict the new data classes and label them in the test phase based on what it had learned. Some categories of the methods of classifications are Decision Trees, Support Vector Machines, Neural Networks and Nearest Neighbors. Some functions of this includes usage in understanding the risk that maybe accompanying a credit, the analysis done to accept a loan request, or to categorize news based on its contents [25].

More information about Decision Trees functions is discussed in section 2.4.7, and more information about Neural Networks is discussed in section 2.4.8.

2.4.3. Association

Association is mostly about linking of a data to another data. For example, buying something may be accompanied by (and linked to) buying something else [22]. This method is somewhat similar to frequent item-sets in data mining that shows items that a customer most frequently purchases together.

2.4.4. Outlier Detection

Outlier Detection tries to find anomalies and can be used in order to prepare for the occurrence of anomalies [22]. Outlier Analysis which is also named anomaly (not normal and not suited cases) detections may use any of data mining techniques for its purpose [25]. Many data mining

techniques tend to ignore anomalies. However, in many cases the anomalies are the focus. For example, in fraud detection [22].

2.4.5. Clustering

Clustering is used to visually show distribution of data [22]. Moreover, clustering is used for grouping big data and high dimension data. It can be done in different methods like hierarchical or point assignment [26]. Clustering is a form of unsupervised learning that does not label data. Instead it focuses on dividing data into different parts based on similarities and differences in a way that more similar data would be organized in the same group [25]. Some models that are introduced in this category would be k-means clustering, fuzzy clustering, hierarchal clustering, and density clustering [25]. This method can be used to study data based on similarities and differences. For example, in marketing the data of customers can be categorized by clustering, in order to help determine the best interaction to attract customers that would be related in the same categories [25].

2.4.6. Regression

This is used to find relationships between variables and is used for forecasting [22]. Regression is somewhat like classification, but it is used more for prediction for continuous data [25]. One model of regression is Linear Regression. Some functional areas of regression usage would be in trend predictions, home prices, and for medical data like human lifespan [25].

2.4.7. Decision Trees

This is used for predictive modeling to see what input has what output and is generally straightforward. However, Decision Trees in combination forms a random forest model, which could be a complex model, and random forest result is not vivid from beginning [22].

2.4.8. Neural networks

Neural networks are kind of machine learning that may be used together with AI and deep learning to do data mining. It employs various layers similar to human neurons and can be one of the most accurate data mining methods. However, how an output is obtained can be very complex to understand [22].

2.4.9. Statistical techniques

Many analyses use different statistical concepts. Some models don't change and are constant, but some others may improve such as in machine learning which uses statistical concepts and weight to recognize things (for example for visual recognition) [22].

2.4.10. Data Warehousing

Pioneer warehousing involved storing data in structured databases to enable retrieval by queries. Newer warehousing involves unstructured or structured databases [22].

2.4.11. Long-Term Memory Processing

This is used to process data gathered over long periods. For example, the data gathered for several years in order to better financial decisions [22].

2.4.12. Machine learning and artificial intelligence

Machine learning and artificial intelligence are usually used for accurate data mining and mostly for computer vision, speech recognition, deep learning, and natural language processing. They are also usually used for semi-structured or unstructured data [22]. More information about AI method in supervised learning can be found in section 2.3.

2.4.13. MapReduce

Distributed systems are made of many low-cost computer components, in order to protect the data from failure. This is achieved by having redundancy of data and parallel data, should a node fail. At the top level of system files, there would be software supporting a MapReduce program that includes Map and reduce functions for parallel processing done on computer nodes. An example of this method is used by Google and Hadoop [26].

2.4.14. Distance Measures

Distance Measures are used to find problems like the nearest neighbors in multidimension environment. They can be used to find similar pages or similar content to find plagiarisms [26].

2.4.15. Data Streaming

Data Streaming needs various streaming managements for updated data that may be available only at a point of time and would be removed forever. It is also used for data that might be available in the future only, or if there is not enough time to process now, it can be first stored and then analyzed later [26].

2.4.16.   Link Analysis

This is used mostly in some search engines to analyze links. For example, to recognize spam from other information in the web. This may be done by following a link or the term used in the webpage [26].

2.4.17.   Social-Network Graphs

This is using graphs to show the social network connections. Social networks are entities such as people having relationship (connection) to each other like being friends such as in Facebook etc. [26].

2.4.18.   Dimensionality reduction

Large data sets can be viewed as huge matrix. In order to make a large matrix smaller, it could be narrowed to narrower matrices that have fewer rows or columns [26]. This process of narrowing matrices is called dimensionality reduction. An example can be Social Network Graphs matrices.

2.4.19.   Frequent Pattern Mining

Frequent Pattern Mining is a method to find previous patterns and interesting patterns in a database. It is not looking to classify or enable clustering and is more related to market analysis, for example for transactions [25].

The mentioned methods of Frequent Pattern Mining in section 2.4.19, Association in section 2.4.3, and Associative Rules in section 2.3.2, all have nearly the same meaning of itemset and dataset that is used in data mining for frequent item set (or data sets) in a transaction. This frequent

item set are mentioned in this thesis section 2.5. According to James Le [26], Frequent Itemset Analysis is like having different items in different baskets. Thus, the goal is to find the frequent item-sets that appear in many baskets.

2.5. Analyzing frequent data

Today's world is a digital world, and we have a lot of data that needs to be processed. Processing of data can lead to a better decision-making. Data mining methods can help extract the main trends and relationships among thousands of records of data. There are various techniques for doing data mining. Two of those techniques are discussed below.

2.5.1. Frequency of Data Sets

According to the information provided by Knowledge Discovery and Data Mining (KDD) [27], frequent pattern data mining finds when items occur together like in a transaction, and these occurrences are tracked to count how often they occur. Subsequently, the number of occurrences, are compared to a defined minimum frequency of occurrence. This is one of the basic concepts of data mining. This concept is named Frequency of Data Sets, and the items happening the most would become the most frequent item-sets. The goal can be to find of the longest frequent item-sets. This method can be used for something like shopping list of customers in each transaction. By doing data mining methods on the list of shopping in each transaction done by customers the most frequent bought items together can be found, the longest list would be called longest frequent data sets.

For example, consider 4 factors X Y Z T and these factors would be items bought together in transactions. Let us assume we have the following occurrences, and we are looking for the

frequency of TYZ, and the minimum accepted frequency to be considered as a frequent itemset needs to be larger than one.

XY

XZ

YZ

XYZT

YXZ

ZYT

TYZ

XY

Considering above transactions, the number of times that TYZ (not sequenced) happens together is 3. While, if sequencing of TYZ occurrence is important, the frequency will be only 1, and so it would be marked as an unfrequented itemset.

2.5.2. Sliding window

Considering the data mining frequency techniques, it is the intention to get results that show the frequency of various "rules' sentences". This includes various factors and rules that show how these factors react to each other. In a case like this thesis where the factors affect each other, all possible factors need to be considered. In order to make data comparable and to see possible trends in the data, the data needs to be divided into different time windows, which can contain one or more events. The time window is gradually stepped over the entire data set, given the name sliding window [28] theory. An example of the sliding window method is show below for a data set of 9 elements, with the window encompassing 3 elements shown in red color.

1 2 3 4 5 6 7 8 9

1 2 3 4 5 6 7 8 9

1 2 3 4 5 6 7 8 9

By using sliding windows, the data can be analyzed part by part, locally, and gradually. There is flexibility of choosing the window size in this method. This method only holds some parts of data each time in a memory, so it reduces the load on memory.

2.6. Data analysis method of this thesis vs. other methods

This thesis presents a new method of data analysis, which was invented several years ago by the author of this thesis (Nadia Khoshgoo). The method consists of a term named Frequent Rules that has some similarities to the frequent itemset in frequent datamining, but its elements have conceptual differences to each item in a datamining itemset (that are usually like the items in a transaction). This new method does not look for classifying, but rather tries to find Frequent Rules that may exist between the data element changes, as well as relationships between them. Comparing the methods and techniques in 2.3 and 2.4, it can be understood that there are some similarities between Association, Associative Rules and Frequent Pattern Mining. Also, these methods have similar definitions to Frequent Item Sets that are discussed in section 2.5.

The bullet points below briefly describe the new method of data analysis which is presented in this thesis, and some of its qualities.

1. In appearance it may look like frequent item sets but, each of its items is different from what is in frequent items sets. Frequent item sets for example is a series of transactions together and frequent one forms an itemset. But in this thesis, frequent rules find the relationships between changes of factors and then forms rules.

2. The window time that is used in this thesis has one main difference compared to the usual definition of window time in data mining, which is that each window does not have overlap

20

with the other windows.

3. This method's main goal is data analysis based on available data and it can form rules that can be interpreted for various areas, and it focuses on the relationships rules between factors vs. each item (factor) occurrence with another.

4. Many data mining methods like neural network are based on predictions and their goal is to learn to enable prediction of new data, and to find mostly arithmetic rules. Considering that a sudden added new variable may affect predictions, a method that does not rely on prediction could be preferable for decision making, especially because predictions can be affected by factors not considered.

5. The method presented in this thesis does not give weight to factors, which is commonly done in some methods of genetic algorithms.

6. Data normalization by deleting outlier data before analysis is not done.

7. In this thesis, the focus has been to find relationships based on classified important information, such as when a factor has been identified to be a Maximum or Minimum.

This new method and the mentioned differences are described in detail in chapter 6 of this thesis.

2.6.1. Why did I invent my method and I did not want to use the other methods?

1. I needed to find relationships of data, but the data analysis and data mining methods that I knew were usually uninteresting for me as they were based on predefined formula (in some cases complex ones), for which I just had to obey the formula and write a code, or they might have had ready pre-written libraries in various programming languages. Therefore,

they did not need any innovation or creativity and honestly that seemed boring for me. So, I found the best way is to make a method based on what I need from the data analysis and data characteristics of this research.

2. The frequent items data analysis method needed items in itemset and transactions to be ready and available, but for a raw data like this thesis, none of these were available. Therefore, in my invented method I made them by using the data relationships and I named them rules and by looking for them I calculated their frequencies.

3. As I wanted relationships between changes, I made the itemset based on the changes.

4. I could not use the usual form of sliding window as it would have cause overlap and would cause double reading of some data, so I changed it to suit the method.

5. I preferred looking for Maximum and Minimum changes based on time rather than daily averages. Therefore, I could find changing relationships and it does not become normalized by averaging, and the relationship would not become lost by normalizing and averaging.

2.7.   Previous work done related and along with this thesis.

In the first phase of this project, Valerie Beynon [2] developed capacitance sensing boards for the hive. These capacitance boards were divided into 12 regions and used capacitance sensors to monitor honey production and bee activity. The data was monitored by an Arduino microprocessor and stored on an SD card. A picture of one of Valerie's capacitance boards is shown below.

The capacitance sensors are made of interdigitated wires on the surface of the frame that goes into the beehive. Each sensor region is 9x7 cm in size. In the thesis of Valerie Beynon [2], various designs were considered for these sensors. The design that was chosen allows for the good detection of changes in capacitance within about 1 cm of the surface of the frame, on both sides,

based on the spacing between the interdigitated wires. As the electric field passes through materials placed over these sensors, the capacitance will change based on the properties and thickness of the material or object. When bees construct honeycomb, lay eggs and raise brood, or when they walk over the frame, the capacitance will then change in relation to the dielectric constant of the material compared to air. For example, the relative dielectric constant of water is 80 (air is 1), and that of honey is about 24 with some variation depending on its water content [2]. Each of these 12 regions of a capacitance senor that is shown in below photo were connected to a separate pin of the Capacitance (Touch) sensor.
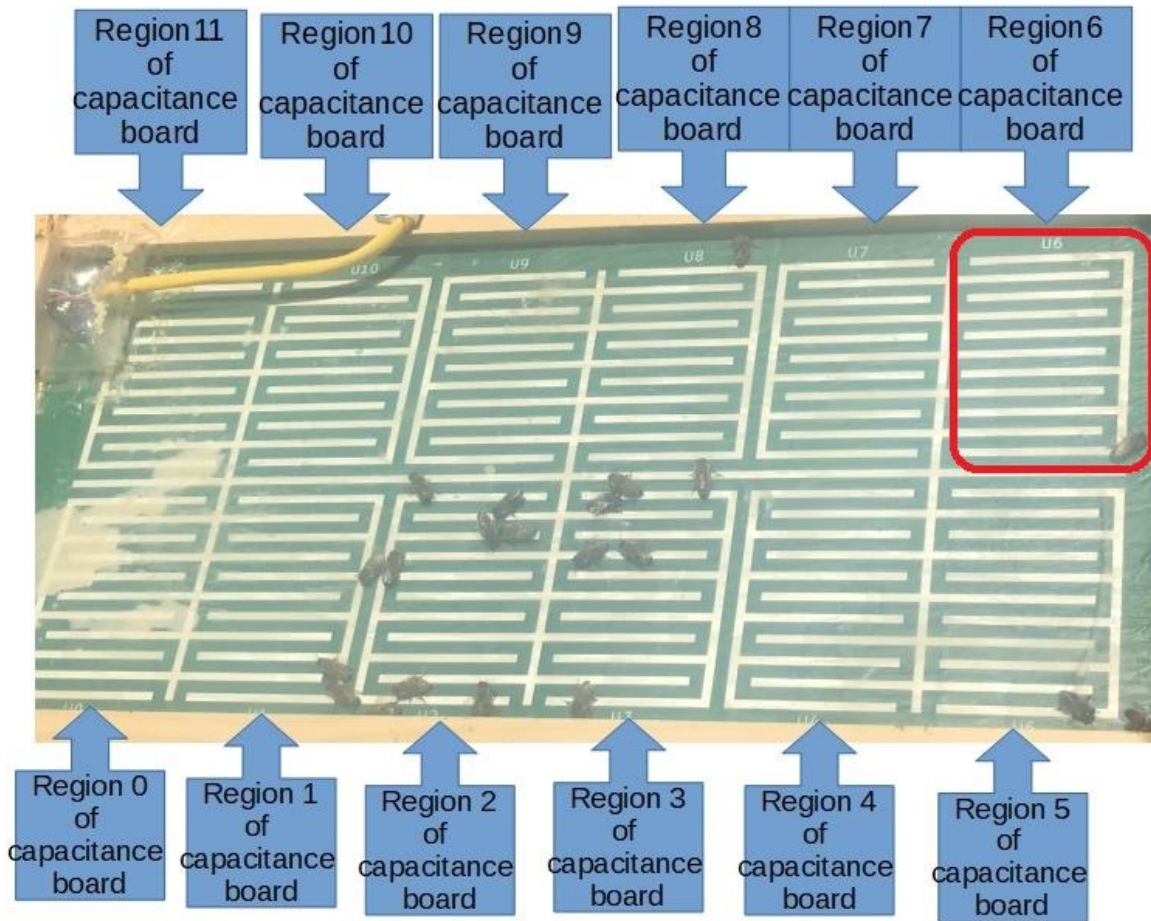
Figure 2.1- Picture of capacitance board designed by Valerie Beynon [2]. The capacitance board is divided into 12 separate regions as labeled in the photo. A sample approximate region is marked by red border in the picture for region 6.

Chapter 3. Phase 1: Implementation of temperature/humidity sensor monitoring system and data analysis software.

As mentioned in Chapter 1, the work of this thesis expanded upon the hive sensor systems developed by Dr. Currie's research group [1] and the capacitance boards of Valerie Beynon [2]. The first phase of this thesis was the implementation of new temperature and humidity sensors to the hive, and development of software to analyze the data obtained from all sensors (new temperature and humidity sensors, Dr. Currie's sensors, and Valerie Beynon's capacitance boards).

Section 3.1 describes a first effort of this project phase which is the implementation of two temperature-humidity sensors inside the hive. This implementation included the design of a data acquisition system to communicate with the temperature and humidity sensors, which used a multiplexer that employed I2C communication. Secondly, the software program was written to read and analyze the recorded data. Section 3.2 describes this software program. This program uses hive data obtained from the experiments of Dr. Currie's research group's web-based system [1] as input.

Section 3.3 describes another software program that is written to analyze data that already had been obtained from Valerie Beynon's prior research using the capacitance boards in the hive [2]. This section also describes the capacitance board hardware.

The goal of this chapter is to discuss the software developed for analysis of the data obtained from the hives. The data is analyzed to give comparison of data, obtained averages, and gradient values of capacitance boards' data. Software programs are developed to read the hive data based on the following common structure. Firstly, each text data file is opened and read once. The file's main data is a combination of text and numbers and are stored in a 1 dimensional array. Also, the

start date and end date of the data are shown in output. If minimum or maximum values, daily averages or gradients need to be calculated for the requested time, they are read and calculated as it is being stored in the array. Moreover, data will not be stored if it us not in the requested period, but memory is reserved for the data in case it is needed. Finally, at the end of software program, only the data from the requested period and some significant points about them are shown in the output. The second and third software programs in this phase are improved versions. These software programs are written in C and C++ languages.

3.1. Temperature/Humidity sensor hardware and software program

This section of the thesis is divided into two parts. Section 3.1.1 is the description of the hardware consisting of two temperature sensors connected by multiplexer to the Arduino and its software programming flowchart.

The Arduino Board [29] is a microcontroller board with input and output pins, including I2C ports, that makes it easy for designers to program and connect peripherals. It has two I2C pins, which are used to connect to sensors employing I2C communication, using SCL and SDA pins. The SCL is the pin sending clock signals and the SDA is the pin for sending receiving data. In section 3.1.2, the data acquisition software program and its flowchart are described. Temperature/Humidity sensors are Adafruit Si7021 [30]. In this sensor type SCL and SDA pins were used to be connected to the multiplexer and Vin and GND pins were connected to the hardware input Voltage and Ground. A tutorial for the temperature/humidity sensor showing how to set it up and how to use it was obtained from Adafruit website [30]. Also, an Adafruit data logger shield board is used [31]. This shield connects the Arduino board to an SD card having timer and the battery slots for the timer. The tutorial of the data logger illustrates how to set up the

data logger board and how to use it, can be accessed from the Adafruit website [31]. This is another board that was used on top of the Arduino board to store the sensors' data in a text file in an SD card.

3.1.1. Temperature/Humidity sensors and Multiplexer-Arduino system

Two temperature/humidity sensors (Adafruit Si7021) [30] were connected to a multiplexer (Adafruit TCA9548A 1-T0-8 I2C Multiplexer) [32], which was then connected to the Arduino [29]. The software program described below was written to enable data to be sent from sensors, via the multiplexer, to the Arduino, which then saves it on an SD card that is inserted in the data logger board [31].

The Arduino has one pair of I2C pins for sending and receiving data and for its clock. The Arduino is connected to the multiplexer board through its one pair of I2C pins (SCL and SDA). The reason a multiplexer was used, was because each temperature/humidity sensors might not be individually addressable via I2C. Therefore, the multiplexer was used to enable the connection of multiple sensors via I2C connection to the Arduino, and with each given the same data address (default address). This was done to simplify the sensor system, and to allow easy replacement of sensors. The multiplexer possess 8 I2C (SDA and SCL pins) which means it can connect up to 8 I2C sensors with same address to the Arduino I2C pins. The tutorial of the multiplexer and how to set it up and use it, was obtained from the Adafruit website. [32].

A software program was written to control the Arduino's function to read the temperature/humidity sensors. The guide for how to write software program and the needed libraries for programming were in the samples in Arduino Application's library and the in the samples in Adafruit website [29,30,31,32]. The software program code has the following sequence.

First, adding the needed libraries to detect and read from the sensors [30], the data logger [31], and multiplexer [32]. After defining the address of the Multiplexer, it will be called to determine if it is connected or not. If it is connected, then the ports of the multiplexer are scanned to find the sensors that might be connected to the multiplexer and to check if they are working or not. After that, the software program reads each sensor one by one and stores its data in a text file on a SD card (also print the output out on the serial monitor output for testing the data in the testing phase). In other words, the software would gather the temperature/humidity data from the sensors in the hive boxes and write them in the files in the SD card. The SD card is located on the data logger board installed on top of the Arduino. The Arduino software program flowchart can be seen Figure 3.1.
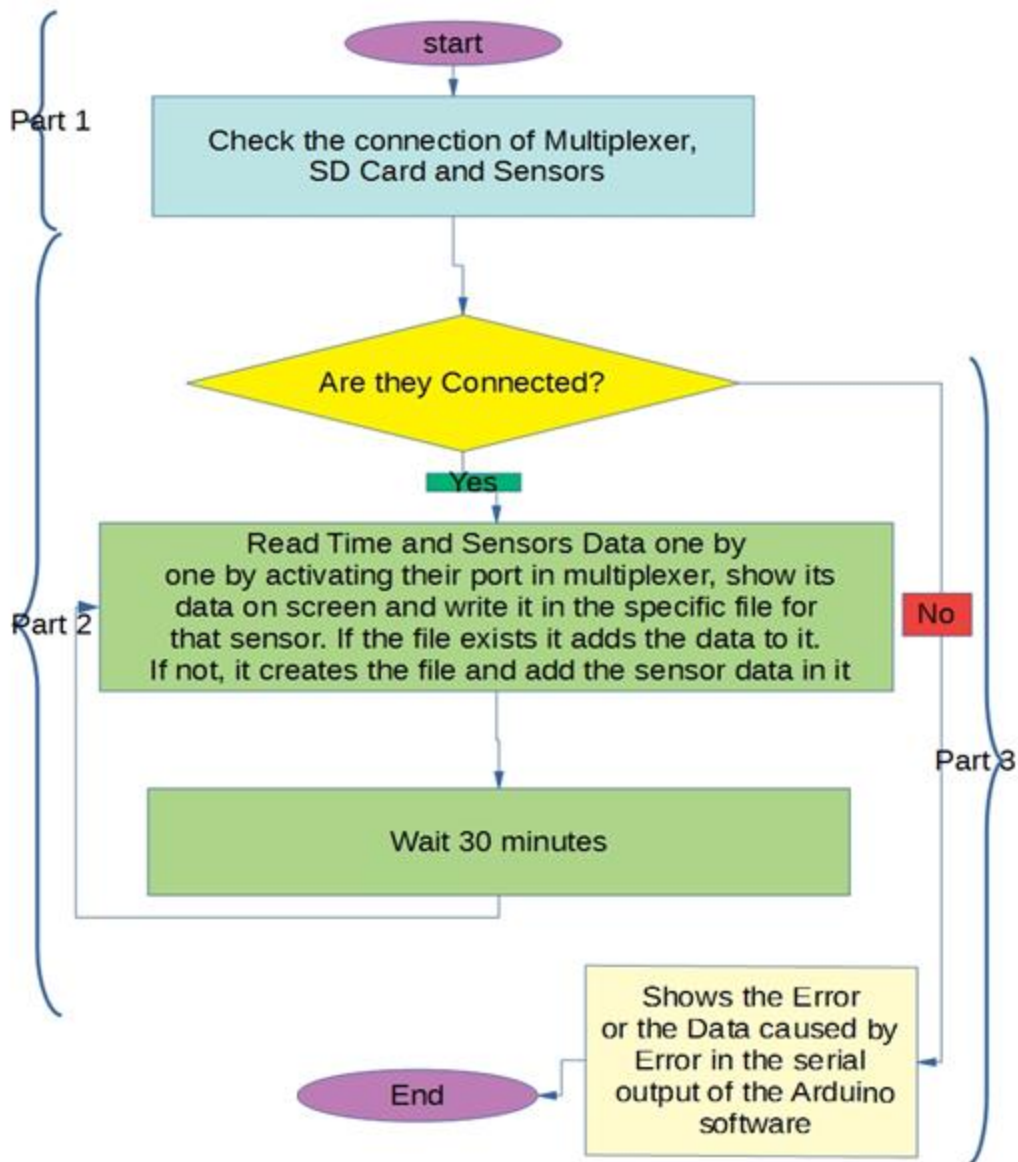
Figure 3.1- Flowchart showing the software program of the Arduino to obtain the data from the temperature/humidity sensors.

The below description describes the sequence in flowchart of Figure 3.2 in more detail. Both, the flowchart in Figure 3.2 and its complementary description in following paragraph are divided into 3 parts.

Part 1)   Initialize the sensors and check their connections to see whether the multiplexer, SD card and sensors are connected and working.

Part 2)   If all electronics are connected and are working correctly, it will read the data from the sensors by selecting the appropriate port of the multiplexer that each sensor is connected to, and send the data to the Arduino. Then it will store the result in the file on the SD card and show the data on the monitor in the serial screen, if it is connected to the computer. Then it waits 30 minutes. Then it will repeat the previous mentioned tasks.

Part 3)   If the connection is lost or the sensors were not detected, it gives an error (or wrong data caused by this bad connection) on the monitor serial screen if it is connected to the computer.

Below is a sample of output by the Arduino software program. Each entry has time information stored, and the saved text is descriptive to make it easier to understand what data is being stored.

2017/10/16 (Monday) 7:7:33, Humidity lowboxfy, 71.04, Temperature lowboxfy, 9.66
2017/10/16 (Monday) 7:38:38, Humidity lowboxfy, 72.92, Temperature lowboxfy, 8.87
2017/10/16 (Monday) 8:9:44, Humidity lowboxfy, 73.05, Temperature lowboxfy, 8.93
2017/10/16 (Monday) 8:40:49, Humidity lowboxfy,71.18, Temperature lowboxfy,8.86
2017/10/16 (Monday) 9:11:56, Humidity lowboxfy,71.43, Temperature lowboxfy,8.80
2017/10/16 (Monday) 9:43:3, Humidity lowboxfy,70.87, Temperature lowboxfy, 8.78
2017/10/16 (Monday) 10:14:11, Humidity lowboxfy,70.35, Temperature lowboxfy,9.21

2017/10/16 (Monday) 10:45:22, Humidity lowboxfy,69.08, Temperature lowboxfy,9.77

2017/10/16 (Monday) 11:16:36, Humidity lowboxfy,66.60, Temperature lowboxfy,11.10

2017/10/16 (Monday) 11:47:51, Humidity lowboxfy,65.32, Temperature lowboxfy,11.94

3.1.2.   Software program for reading Arduino output data file

In this section, the software program that reads and stores the data obtained from the temperature/humidity sensors is described. The data is stored in two separate text files, one is named "Lowbox" and the other is named "Top box". The data in these text files is preceded by information related sensor description. However, these descriptions were redundant, and they do not need to be stored in the software program, so they were removed later. This software extracts the valuable data (the numbers and the date) that are stored in two text files and shows their amount on the screen. The whole data is stored in three separate 1-dimensional arrays. After reading the data of each file, (the upper and lower box data) separately, their data is printed on the output screen, so it shows that their data is stored correctly. The below description describes the steps, and this is shown in the flowchart of Figure 3.2.

Part 1)   Checks if the file exists and does it have data and whether it opens it.

Part 2)   According to the format of the temperature data, remove unimportant data like the title before sensor data. After finding the data, only the numeric amount is stored in arrays. To check whether the software program works well, the output shows the input data including date and time followed by sensor's data.

Part 3)   If the file does not exist or no data is available, it will give an error message. Also, if the data has finished, it will show the end of data message.
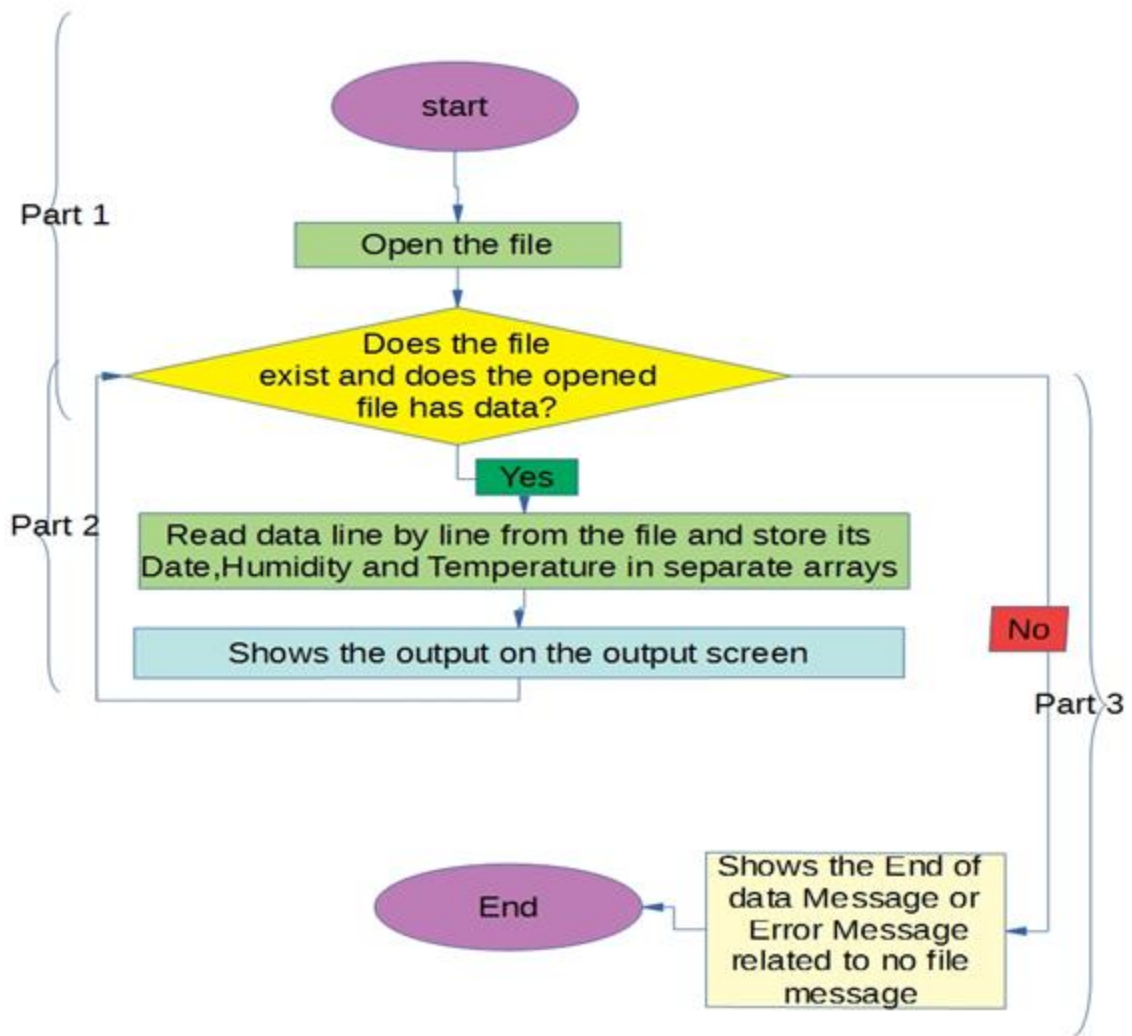
Figure 3.2- Flowchart of the software program for reading for temperature sensors and multiplexer hardware hive.

3.2. Software for interpreting data from online hive monitoring system

This section describes the software that was written for performing data analysis on the data file obtained from Dr. Curry's research group [1]. Dr. Curry's data include information from the following sensor types: outdoor humidity and temperature, and hive weight scale. This data was collected by a web-based system.

There are some points that needed to be addressed in the software. The original data from Dr. Curry's group was a large file, having a lot of zeros for some time periods. The zeros were for periods that perhaps their system was not installed and operated fully. To fix this issue and for using as an input for a software, the zero containing records were deleted by using Macro scripts, in Excel [33]. Then, the columns of needed data were extracted to a text file and a software was written to extract the useful data from it. Figure 3.3 shows a flowchart, and a description of how the software works is given below.

Part 1) The user is asked to "Enter the Start Date" and "Enter the End Date" intervals which the user wants to obtain its maximum and minimum amount and Daily fluctuation. Also, it asks whether the user wants the daily differences amount other than zero. In other words, if the user wants the fluctuation base amount to be larger than zero, the user enters their chosen amount.

Part 2) Moreover, it tries to open the file. In case it exists, it starts reading from the file line by line.

Part 3) It calculates the minimum and maximum of each data type and stores the data. This is calculated while the reading, so the file is read only once.

Part 4)   If it is between the start date and end date, it calculates the daily averages and its fluctuation based on the Fluctuation base. This is calculated while the reading, so the file is read only once. The output is shown on the screen.

Part 5)   If the file does not exist, it will show the message error. Moreover, it will show end of file data when the file reading has finished.

The software program asks the user to enter start and end date, and asks for the minimum fluctuation amount, if other than 0. It opens the data file, extracts the useful information, and stores the variables' amount and find minimum and maximum of values for the weight, humidity and temperature. If the entered date is between the start date and end date, it calculates the daily averages for all sensors. The software determines if daily average is increased or decreased from the value of the previous day's average.
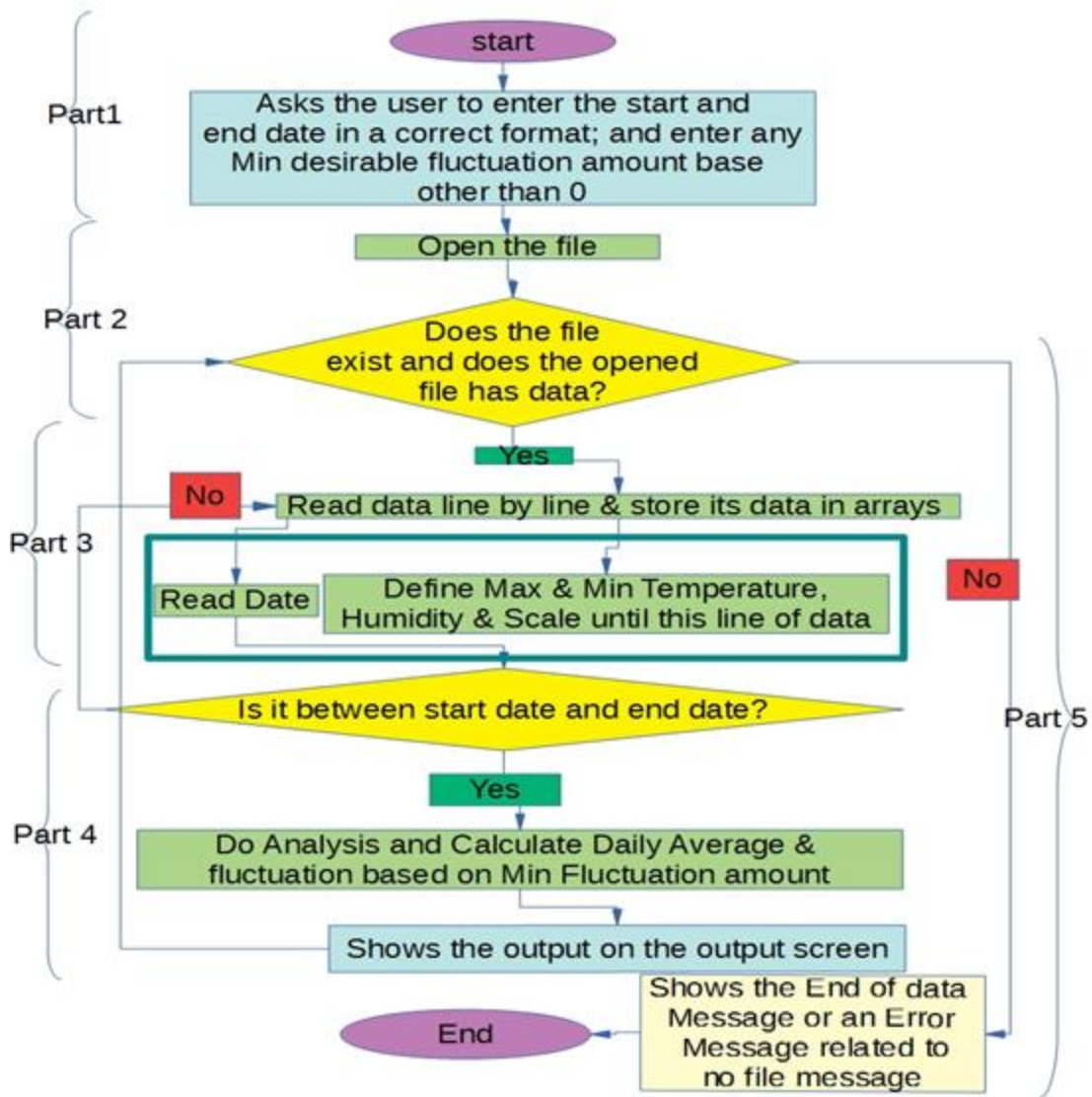
Figure 3.3- Flowchart of software programming for reading data from Dr. Curry's research group [1].

3.3. Software for interpreting capacitance board data

In this section the software for gathering data from the capacitance boards developed by Valerie Beynon in her prior research in 2017 [2]. This capacitance board is divided into 12 regions and used capacitance sensors to monitor honey production and bee activity. The capacitance board is described in section 2.4.

Software was written that reads the data in each capacitance region 10 times every half an hour and records the average of the 10 readings in the data file. The software then determines the slope between two consecutive time periods for the following time periods: two nights, two days, mid-day, and midnight. The "Daily averages" is also calculated to show the overall daily activity of bees on each part of the capacitance board being investigated. At the same time the slope would help to compare a data to the previous day's value, the overall trend for two consecutive days can be seen. The software also gives the flexibility to the user to choose the period that these calculations will be reported for. This software is described in the flowchart of Figure 3.4, and a description is given below.

Part 1) It asks the user to enter the Start Date and End Date Interval that the user want to analyze its Daily Changes.

Part 2) It tries to open the file. In case it exists, it starts reading the file line by line.

Part 3) If data time is between the start date and end date, the software averages the ten reading for each point of the time then it calculates the daily averages and Noon and Midnight daily changes to its previous Noon and Midnight slope and shows its result by a visualization shape at the end. This is calculated during the reading, so the file is read only once. The output is shown on the screen.

Part 4)   If the file does not exist, it will show the message error. Also, it will show end of file data

when the file reading has finished.
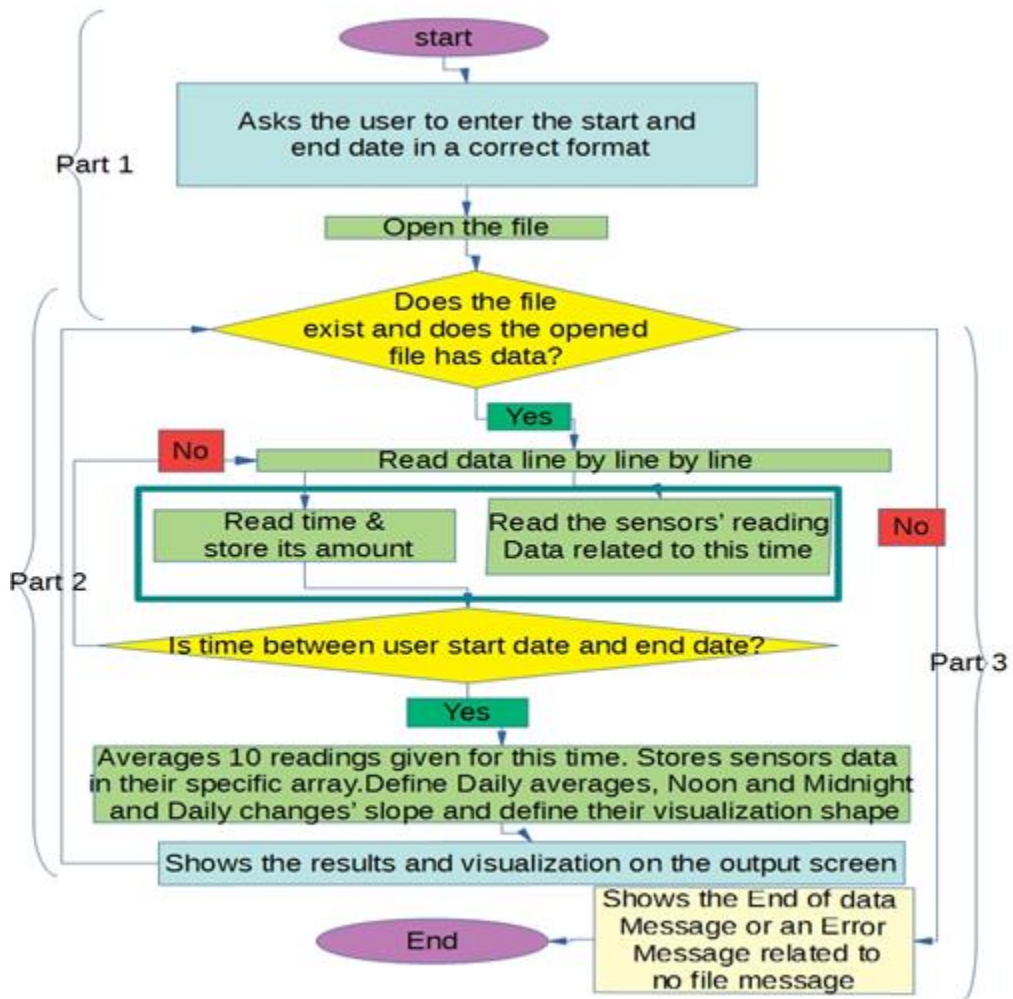
Figure 3.4- Flowchart for software that used the input of the capacitance board data. The board was made by Valerie Beynon [2]

The format of the stored data from the software is shown below:

 day #, Time of day, Actual Date
 sensor 11, sensor 10, sensor 9, sensor 8, sensor 7, sensor 6
 sensor 0, sensor 1, sensor 2, sensor 3, sensor 4, sensor 5

Below is a sample output of this software:

 IntvdaysNumber: 3 , Noon Amount, 09/10/2017
 16.86, 18.49, 20.29, 22.42, 24.01, 25.81
 18.16, 22.06, 23.58, 24.46, 23.46, 25.32

 IntvdaysNumber: 3 ,Noon Amount to yesterday amount slope 09/10/2017
 0.0, 0.06, 0.06, 0.09, 0.09, 0.11
 0.05, 0.09, 0.08, 0.11, 0.23, 0.16

 IntvdaysNumber : 3 ,Midnight Amount of 09/11/2017 :
 16.95, 18.54, 20.33, 22.48, 24.06, 25.87
 18.24, 22.08, 23.63, 24.58, 23.53, 25.41

 IntvdaysNumber: 3 , Midnight Amount to yesterday amount slope 09/11/2017
 0.12, 0.07, 0.05, 0.08, 0.08, 0.08
 0.11, 0.26, 0.06, 0.15, 0.19, 0.04

 IntvdaysNumber 3, Daily Averages of 09/10/2017 :
 16.87, 18.50, 20.31, 22.42, 24.02, 25.83
 18.21, 22.05, 23.62, 24.49, 23.40, 25.34

To help the user understand the data, software was written to help visualize software program's output. "Interval day's number" start from 1 and it begins to calculate mentioned parameters from starting date user interred. In below visualization each '$' sign is assigned to a number greater than 0 and equal less than 1. As an example, in the "Average day" box output in below output, 22 number of '$' means a number greater than 21 and equal and less than 22. In this way 21.31 amount is shown by 22 number of '$' in "Average day" box, below. As the slope to previous day does not exist in the first day of interval (as this is the first day chosen by user) there is no '$' in the slope box of interval day number one showing zero amount.
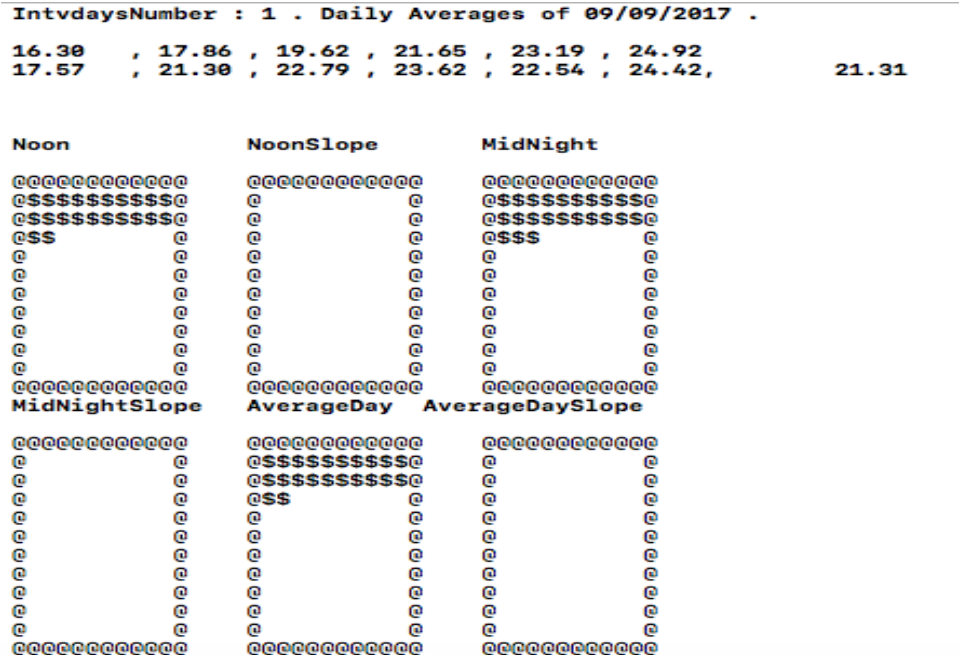
```
IntvdaysNumber : 1 . Daily Averages of 09/09/2017 .

16.30    , 17.86 , 19.62 , 21.65 , 23.19 , 24.92
17.57    , 21.30 , 22.79 , 23.62 , 22.54 , 24.42,            21.31


Noon                   NoonSlope              MidNight

@@@@@@@@@@@@           @@@@@@@@@@@@           @@@@@@@@@@@@
@$$$$$$$$$$$@          @            @         @$$$$$$$$$$$@
@$$$$$$$$$$$@          @            @         @$$$$$$$$$$$@
@$$$       @          @            @         @$$$        @
@          @          @            @         @           @
@          @          @            @         @           @
@          @          @            @         @           @
@          @          @            @         @           @
@          @          @            @         @           @
@          @          @            @         @           @
@@@@@@@@@@@@           @@@@@@@@@@@@           @@@@@@@@@@@@
MidNightSlope          AverageDay   AverageDaySlope

@@@@@@@@@@@@           @@@@@@@@@@@@           @@@@@@@@@@@@
@          @          @$$$$$$$$$$$@          @           @
@          @          @$$$$$$$$$$$@          @           @
@          @          @$$$        @          @           @
@          @          @           @          @           @
@          @          @           @          @           @
@          @          @           @          @           @
@          @          @           @          @           @
@          @          @           @          @           @
@          @          @           @          @           @
@          @          @           @          @           @
@@@@@@@@@@@@           @@@@@@@@@@@@           @@@@@@@@@@@@
```

In the same way the output for the next day is shown below. As the output is about the interval day number two, the slope has become more than zero and less than 1 (0.78), and so it is shown by one '$' in "AveragedaySlopes" box below.

```
IntvdaysNumber : 2 . Daily Slope of 09/10/2017 to yesterday

0.58     , 0.63 , 0.69 , 0.77 , 0.84 , 0.92
0.63     ,  0.75 , 0.84 , 0.86 , 0.86 , 0.93       0.78


Noon                   NoonSlope              MidNight

@@@@@@@@@@@@           @@@@@@@@@@@@           @@@@@@@@@@@@
@$$$$$$$$$$$@          @$          @          @$$$$$$$$$$$@
@$$$$$$$$$$$@          @            @         @$$$$$$$$$$$@
@$$$       @          @            @         @$$$        @
@          @          @            @         @           @
@          @          @            @         @           @
@          @          @            @         @           @
@          @          @            @         @           @
@          @          @            @         @           @
@          @          @            @         @           @
@@@@@@@@@@@@           @@@@@@@@@@@@           @@@@@@@@@@@@
MidNightSlope          AverageDay   AverageDaySlope

@@@@@@@@@@@@           @@@@@@@@@@@@           @@@@@@@@@@@@
@$         @          @$$$$$$$$$$$@          @$          @
@          @          @$$$$$$$$$$$@          @           @
@          @          @$$$        @          @           @
@          @          @           @          @           @
@          @          @           @          @           @
@          @          @           @          @           @
@          @          @           @          @           @
@          @          @           @          @           @
@          @          @           @          @           @
@@@@@@@@@@@@           @@@@@@@@@@@@           @@@@@@@@@@@@
```
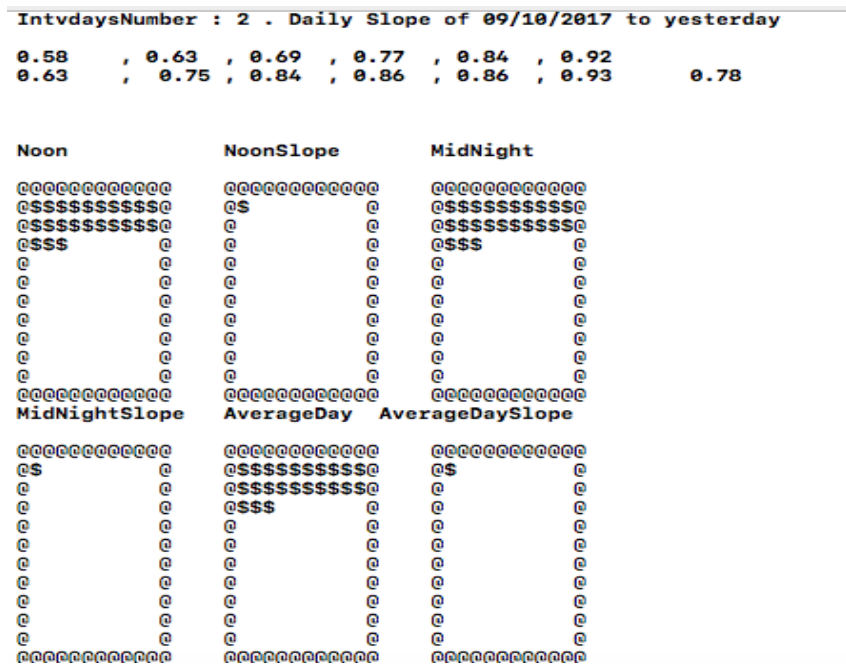
It should be mentioned that if the user's selected exact start date and time does not exist, the software will show the available data from the next available start date. Then, it prints the output result in the output screen; and shows the amount of the Midnight, Noon amount, and averages and slopes of requested period by a graphic like shape. Also, it averages the whole capacitance amount on the capacitance board and writes it in the right side of the 12 sensors daily averages or slopes. In the above examples, this value is 21.31 for Interval Day 1 Average, and 0.78 for Interval Day 2 Slope.

3.4. Conclusion of Chapter 3

In this chapter, the software that was written to read the data from the capacitance boards and sensors of Dr. Curry's group was described. The software does some data analysis to determine the daily averages and slopes of the data, for an interval period entered by user. Also, a visual demonstration is presented that can be a tool for a quick comparison of data by the user. This visualization helps the user review all details of capacitance board's activity average and slope with a quick glance.

Chapter 4. Phase 2: New hardware implementation with expanded sensors

This chapter, and chapters 5 and 6, represent phase 2 of this thesis. An expanded sensor set was implemented in a hive, consisting of 1 brood box and 1 honey super. It included a light sensor outside the hive (to measure outdoor light level), temperature/humidity sensors inside the two hive boxes, and 6 capacitance boards (3 in the brood box and 3 in the honey super). To enable this larger sensor system, the Arduino and multiplexer interface system (that was described in section 3.1) was used as the basis for the new hardware implementation. The multiplexer system enabled the connection of these 9 sensors to the Arduino using I2C communication. This expanded system was developed by Valerie Beynon, based on the multiplexer software of this thesis (section 3.1). In this chapter, the expanded sensor system (hardware and software) is described. In Chapters 6 and 7, the software developed to analyze the data from these sensors will be discussed in detail.

4.1.   Expanded sensor system

The expanded sensor system (Adafruit APDS9960 Digital RGB Light Sensor [34], 2 temperature/humidity sensors [30], 6 capacitance boards using MPR121 microchip [35]) were all connected to the Arduino with two multiplexers. Capacitance boards 1, 2, and 3 were put in the brood box from left to right respectively. And Capacitance boards 4, 5, and 6 were put in the honey super box from left to right respectively. The multiplexers enabled the addressing and activation of one sensor at a time to read data [32]. On the Arduino board, the address of the multiplexer is assigned by its three address ports. As a result, there will be the capability of total number of 2^3=8 multiplexers working together. In this project only two multiplexers were used together, since each multiplexer has eight I2C ports. Capacitance sensor reading electronics are Adafruit MPR121 12-

Key capacitance Touch Sensor Breakout [35] having SCL and SDA pins. These chips are commonly used to read capacitive touch sensors, for implementations having 12 touch pins sensing 12 different areas or object (such as a keypad) [35]. In other words, if it is touched, the number that it sends back to Arduino by software programming changes which is showing that it has been touched. The tutorial of the capacitance sensor, how to set it up and how to use it was obtained from the Adafruit website [35].

4.2. The Arduino / multiplexer sensor connection system

This hardware consisted of Arduino Board [29], data logger shield board [31], and Multiplexer (Adafruit TCA9548A [32]), which some description about these hardware usages in this thesis is in Glossary section of this thesis. The tutorials, guides of individual multiplexer and sensors and their Arduino software program tests and libraries are available at the mentioned Adafruit website.

The sensors were connected to two multiplexers (Adafruit TCA9548A) [32], which was then connected to the Arduino [29]. Each multiplexer is addressable, so it is being called by its address being activated to read its data, then it would become deactivated to read next multiplexer. The data is sent from sensors to the multiplexer and then from multiplexer to Arduino and is saved in a SD card which was inserted in a data logger board [31]. The guide for how to write software program and the needed libraries and guide about the Arduino board were in the Arduino library samples, Adafruit website samples and Arduino Data sheet respectively [29,31].

The software program code has the following sequence. First, adding the needed libraries to detect and read from the sensors, the data logger [31], and multiplexer [32]. After, defining the address of the multiplexer it will be called to determine if it is connected or not. If it is connected, then the ports of the multiplexer are scanned to find the sensors that are connected to the

multiplexer and to check if they are working or not. After that, software program reads each sensor

one by one and stores its data in a text file on a SD card (also print the output out on the serial

monitor output for testing the data in the testing phase).
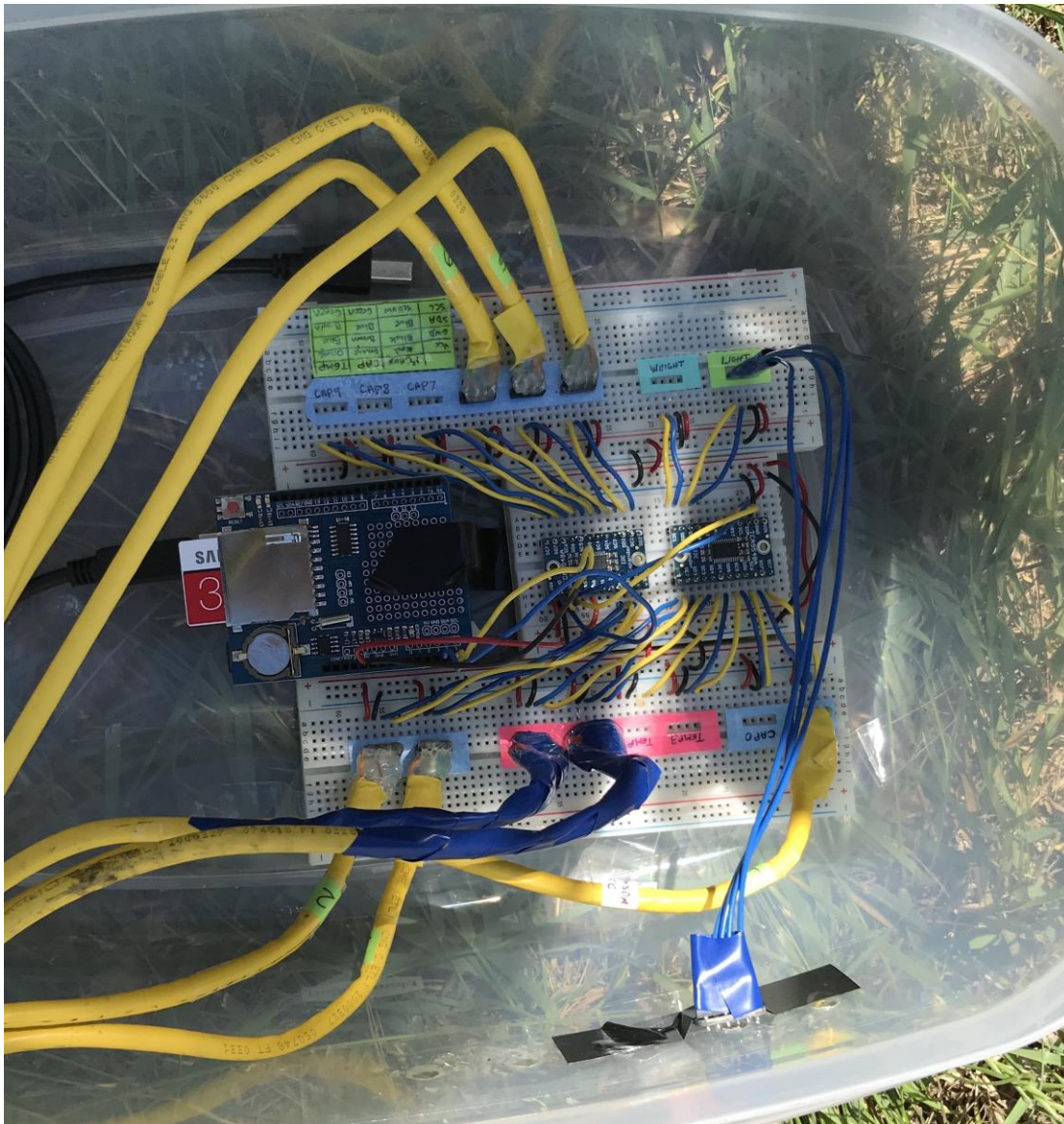
Figure 4.1- Expanded sensor data acquisition system implemented by Valerie Beynon. Photo taken by Dr. Cyrus Shafai (Photo usage permission for this thesis is given from Dr. Cyrus Shafai).

4.3.   Description of acquired data

The sensor data recorded by the Arduino was stored by the data logger shield board on an SD card using the following format. All data is stored as numerical values, with each sensor's data on a separate line in the data file. The system and its Arduino software were implemented and written by Valerie Beynon [2] and the system can connect 1 light sensor, 1 weight sensor, 4 temperature/humidity sensors, and 10 capacitance boards (each with 12 regions of measurement). The below paragraph describes a single hardware reading.


```
Date:       Month/day/Year, Time
Light:      red, Green, blue, Clear, Lux
Weight:     W
Temperature/humidity 1:     T, H
Temperature/humidity 2:     T, H
Temperature/humidity 3:     T, H
Temperature/humidity 4:     T, H
Board 0:    C1, C2, C3, C3, C4, C5, C6, C7, C7, C8, C9, C10, C11, C12
Board 1:    C1, C2, C3, C3, C4, C5, C6, C7, C7, C8, C9, C10, C11, C12
Board 2:    C1, C2, C3, C3, C4, C5, C6, C7, C7, C8, C9, C10, C11, C12
Board 3:    C1, C2, C3, C3, C4, C5, C6, C7, C7, C8, C9, C10, C11, C12
Board 4:    C1, C2, C3, C3, C4, C5, C6, C7, C7, C8, C9, C10, C11, C12
Board 5:    C1, C2, C3, C3, C4, C5, C6, C7, C7, C8, C9, C10, C11, C12
Board 6:    C1, C2, C3, C3, C4, C5, C6, C7, C7, C8, C9, C10, C11, C12
Board 7:    C1, C2, C3, C3, C4, C5, C6, C7, C7, C8, C9, C10, C11, C12
Board 8:    C1, C2, C3, C3, C4, C5, C6, C7, C7, C8, C9, C10, C11, C12
Board 9:    C1, C2, C3, C3, C4, C5, C6, C7, C7, C8, C9, C10, C11, C12
```


Figure 4.2 is one time reading of data with a description added for each line data. Each data line has a description is added in front of it (shown by the blue arrows in the Figure 4.2).
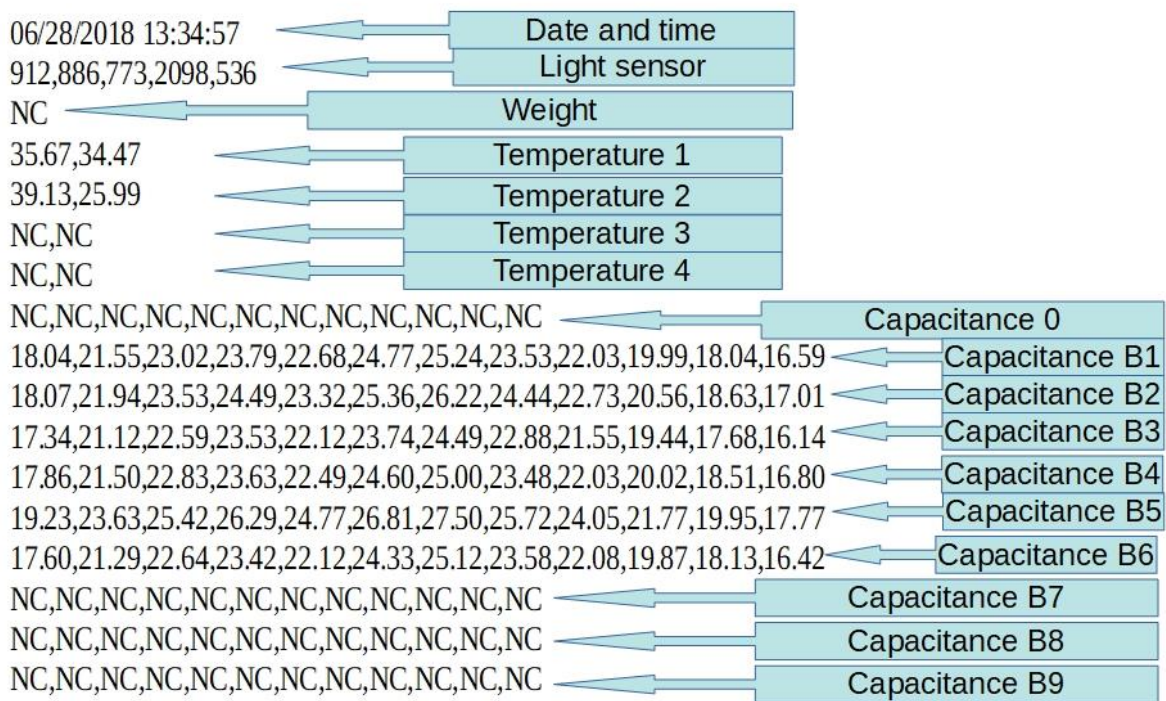
Figure 4.2- One-time 2018 sensors reading. Arduino Program by Valerie Beynon. Original output data from Valerie Beynon [2] prior research. Data description image marked and edited descriptive image by Nadia Khoshgoo.

In nearly all uses for this hive system data file gathering study, a complete set of sensor data was gathered at 30 minutes intervals. Exceptions are for test cases, when the system's battery was disconnected, the system was restarted, or some of the sensors were disconnected/connected to the Arduino.

4.4.   Chapter 4 Summary

This chapter showed the system hardware and the format of the data file output by the Arduino program. Some illustration about the hardware parts is found in the resources cited in this chapter and can be found in the Glossary.

Chapter 5. Visual inspection of capacitance boards for in-hive studies of the Phase 2 hardware

In this chapter, the photos show the implementation of the hive monitoring system hardware described in chapter 4, during 2018. The hive boxes were located at the University of Manitoba Campus under supervision of Dr. Curry (Department of Entomology).

5.1. Placing the capacitance boards in the hive boxes

The capacitance boards were first placed in the hive on June 28, 2018. The hive consisted of two boxes, a lower brood box, and an upper honey super. Six capacitance boards were placed in the hives, with capacitance boards 1-3 in the brood box and capacitance boards 4-6 in the honey super.

Board 2 already had some combs attached from the prior year's experiments. Figure 5.1 and Figure 5.2 show the capacitance boards made by Beynon [2]. The locations of the capacitance boards in the hive boxes are described in the figure's captions.

Figure 5.1- Brood box. Capacitance board 1 is on the left, capacitance board 2 in the middle, capacitance board 3 on the right.

Figure 5.2- Honey super box. Capacitance board 4 is on the left, capacitance board 5 in the middle, capacitance board 6 on the right.

The below timeline shows the main points observed while inspecting the hive:

- The bees started building combs on capacitance board 4 and capacitance board 1 inner sides (right side) around July 11, 2018.

- The bees started building combs on the capacitance board 5 around July 18, 2018.

- The rate of comb building on capacitance board 5 that was observed was increased by the July 25, 2018 hive inspection.

- By August 1, 2018 capacitance board 5 had more comb on it than capacitance board 4.

- By August 1, 2018 the bees had built more comb on the left side of capacitance board 2.

- The bees seem to not like the right side of the hive. It maybe because this side of the hive is under sun more than the left side, because of the trees making shade. This shade could act like a shelter from the extreme sun exposure especially in summertime.

5.1.1. Adding wax starter frames to capacitance boards

This section describes the changes done to increase the bee's activity on the right side of the hive. Several photos of the hive capacitance boards are also shown.

On August 15, 2018, in order to accelerate comb building, pre-prepared thin comb wax starter frames were added to each of the capacitance boards, as shown in Figure 5.3.

Figure 5.3- Wax starter frame added to capacitance board 3. Photo taken 15 August 2018.

To verify that the capacitance board still works with the wax comb on it, I put my hand on the capacitance board, and the gathered data showed that it sensed my hand that was touching the capacitance board, indicating that it was still working with the wax starter frame. The data changes are small and are shown in Table 5.1.

Table 5.1. Capacitance board data reading my hand, before and after putting wax starter frame on the capacitance board.

Capacitance sensor	State		
	Before Putting wax	**After putting wax with my hand on part of it**	**After putting wax and after I removed my hand**
capacitance1	17.43 pF	19.17 pF	18.29 pF
capacitance2	21.55 pF	22.88 pF	22.26 pF
capacitance3	22.78 pF	24.16 pF	23.79 pF
capacitance4	23.95 pF	24.89 pF	24.60 pF
capacitance5	22.68 pF	24.05 pF	23.53 pF
capacitance6	24.27 pF	26.81 pF	25.60 pF
capacitance7	24.89 pF	27.02 pF	26.22 pF
capacitance8	23.79 pF	24.38 pF	24.33 pF
capacitance9	22.45 pF	23.12 pF	22.83 pF
capacitance10	19.91 pF	21.55 pF	20.80 pF
capacitance11	18.22 pF	19.91 pF	19.30 pF
capacitance12	16.42 pF	17.51 pF	16.99 pF

Table 5.2 shows the reading differences sensed after putting wax on capacitance board 3. It shows that the touch sensor was still working well after putting wax on the board, and that the baseline capacitance reading of capacitance board 3's with wax on it does not differ much with and without wax.

Table 5.2. Capacitance board data differences reading my hand, before and after putting wax starter frame on the capacitance board.

Capacitance sensor	State		
	My hand on the capacitance board, difference before and after putting wax	Wax starter frame, difference between my hand on capacitance board and removed	Difference between my hand on the capacitance board before wax, and after wax without my hand on the board
capacitance1	1.74 pF	0.88 pF	0.86 pF
capacitance2	1.33 pF	0.62 pF	0.71 pF
capacitance3	1.38 pF	0.37 pF	1.01 pF
capacitance4	0.94 pF	0.29 pF	0.65 pF
capacitance5	1.37 pF	0.52 pF	0.85 pF
capacitance6	2.54 pF	1.21 pF	1.33 pF
capacitance7	2.13 pF	0.8 pF	1.33 pF
capacitance8	0.59 pF	0.05 pF	0.54 pF
capacitance9	0.67 pF	0.29 pF	0.38 pF
capacitance10	1.64 pF	0.75 pF	0.89 pF
capacitance11	1.69 pF	0.61 pF	1.08 pF
capacitance12	1.09 pF	0.52 pF	0.57 pF

5.1.2.   Results of August 22, 2018 inspection, after adding wax starter comb to the

capacitance boards

Figures from 5.4 to 5.9 show the state of the capacitance board in the hives at the inspection of August 22, 2018. It can be seen that:

- There was a little more comb on left side (towards box corner) of capacitance board 2.

- There is still no additional comb made on capacitance board 3 by the bees. It also appears that the bees have consumed some of the wax.

Figure 5.4- Capacitance board 1 frontside (right), backside (left). Picture taken August 22, 2018.

Figure 5.5- Capacitance board 2 frontside (right), backside (left). Picture taken August 22, 2018.

Figure 5.6- Capacitance board 3 frontside (right), backside (left). Picture taken August 22, 2018.

Figure 5.7- Capacitance board 4 frontside (right), backside (left). Picture taken August 22, 2018.

Figure 5.8- Capacitance board 5 frontside (right), backside (left). Picture taken August 22, 2018.

Figure 5.9- Capacitance board 6 frontside (right), backside (left). Picture taken August 22, 2018.

Moving a capacitance board to test bees' reaction

On August 22, 2018 capacitance board 3 was moved to different location in the hive. The new spot is shown in Figure 5.10.

- When the capacitance board was removed completely from the hive on October 3rd it was found that capacitance board 3's change of location to the center of the box did cause a little increase in bees making honey on it.

Figure 5.10- New place of capacitance board 3. Photo taken August 22, 2018.

Disconnecting the Brood box that contained the capacitance boards 1, 2 and 3

To test the assumption that maybe the bees would not like to be very active on a capacitance board that is powered on all the time, the capacitance boards in the brood box (boards 1, 2, and 3) were disconnected from the Arduino board, so we would test it at the end to see whether this would result in the bees building more comb on them.

- When the capacitance boards were removed from hive on October 3, it was found that the disconnecting of the capacitance boards did not have an apparent effect on the bees making honey on them. Therefore, it could be that having the electrical power continuously connected to the capacitance boards does not seem to be a problem for the bees.

Date error in the stored data after August 22

The 22nd of August was the last date of the correct date time gathered from the hive. This is because after disconnecting some capacitance boards from the hive on August 22, the date of the stored data became reset to a wrong date. Accordingly, in chapters 6 and 7, the data analyzed from the capacitance boards is until the 22 of August 2018.

5.1.3. Last visit of hive photos

On the 3 of October 2018 the capacitance boards were removed from the hive. They are shown in figures 5.11 to 5.15. We observed the following:

- The bees did not build any comb on one side of capacitance board 1, capacitance board 4 and both sides of capacitance board 6, so their photos are not in the following pages.

- The most honey was on capacitance board5 on both sides.

- One side of capacitance board4 had honey.

- Board 1, 2 and 3 had little honey them.

- The number of bees on capacitance board 1 is greater on the side facing capacitance board 2.

Figure 5.11- Capacitance board 1 frontside (right), backside (left). Picture taken October 3, 2018.

Figure 5.12- Capacitance board 2 frontside (right), backside (left). Picture taken October 3, 2018.

Figure 5.13- Capacitance board 3 frontside (right), backside (left). Picture taken October 3, 2018.

Figure 5.14- Capacitance board 4. Picture taken October 3, 2018.
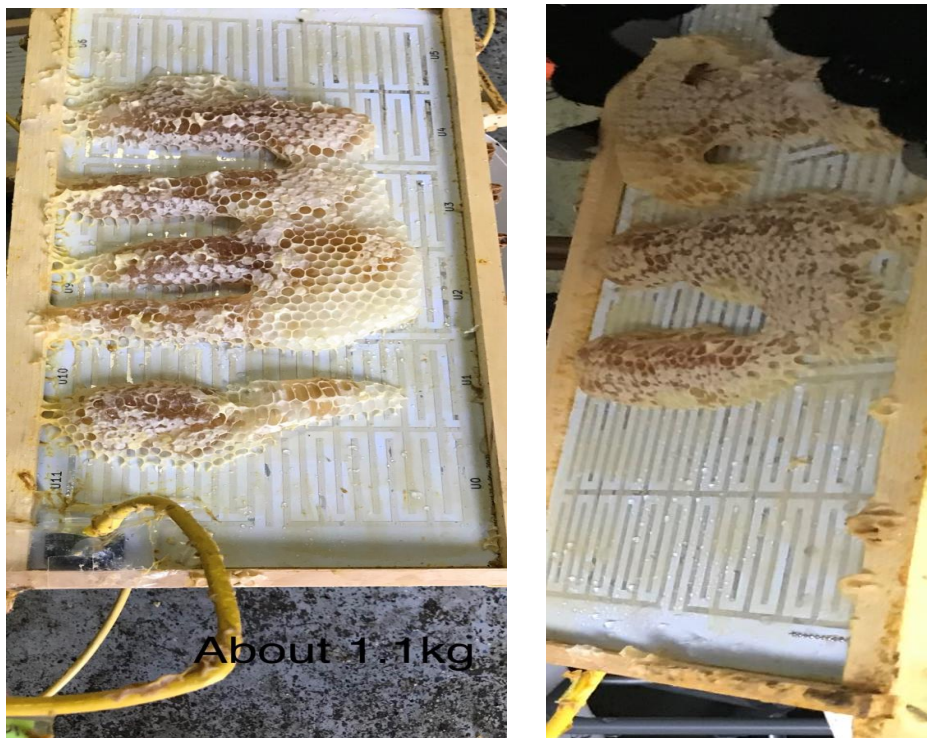
Figure 5.15- Board 5 frontside (right), backside (left). Picture taken October 3, 2018.

Last day Visit Results and Summary

The results of the last day visit and a summary of findings are discussed in the bullet points below.

- Considering the effect of disconnecting the capacitance boards from power, it did not help to cause the bees to build more comb on the capacitance boards in the brood box (Boards 1, 2, and 3). It may be because it was late in the fall and early winter and the weather was too cold for making more comb or maybe the fact of connection of the capacitance board to the electricity does not affect comb building.

- Moving capacitance board 3 to the middle of the box resulted in bees producing some honey on it. However, the bees did remove a lot of the wax starter comb.

- After this day, considering the cold weather and snowy weather, it is understandable that they could not make more amount of honey on the capacitance board.

- At the end of the summer, capacitance boards 4 and 5 had the most honey on them. Capacitance board 5 had a lot of honey on both sides, while only one side of capacitance board 4 had honey.

- Boards 2 and 5, and one side of capacitance board 1 had the greatest number of bees on them during the visits.

Chapter 6. Data analysis to obtain the frequent longest rules amongst sensor datasets - Phase 2 of thesis project

This chapter discusses the software developed for data analysis for the sensor data obtained in 2018 using the expanded sensor system described in chapter 4. First, the software reads the data obtained from the data file stored by the Arduino and organizes it to fit in a 3D array. Then various data analysis methods are explored. This chapter discusses various topics regarding the analysis of data collected by the sensor hardware. These methods compare data and find possible relationships between data or groups of data. Three different methods were developed to compare data relationships. They are:

- **Comparing two sensor data (section 6.2):** This method calculates averages or slope of sensor data, which is then used for comparing these data from two sensors.

- **Longest All-Factors-Relationship (section 6.3 and 6.5):** This method is used to find a relationship between the data of several sensors. The relationship is designated as follows. First, a given sensor data is swept for situations where it has a maximum or minimum value, within a given time window. Second, the other sensor's data is checked to see if its value has increased (positive change - P), decreased (negative change - N), or remained constant (S). This is repeated for all sensors, and their individual results are attached together, to define a "rule sentence". Finally, the analysis explores how often these rule sentences occur (the frequency). The appearance of this rule sentences has similarities to frequent itemset terms in data mining, which is described in section 2.5. However, considering what each letter in factors rules' sentence represents, there are main differences

between the items in itemset in data mining and factors rules' terms (S, N, P). This will be described in 6.6.

- **Two-Factors-Relationship (Chapter 7)**: In this method, one sensor dataset is studied to determine possible relationships with the datasets from other sensors. First, the sensor under study is determined for when it is either maximum or minimum. Second, the other sensors are investigated for the frequencies of positive changes (P) and negatives changes (N). As an example, when the light sensor reads a maximum, the frequency of when the temperature is increasing (positive - P) and decreasing (negative - N) is determined.

The **Longest All-Factors-Relationship and Two-Factors-Relationship** methods both divide the data into small time windows, in order to compare them. This is similar to the sliding window method (explained in section 2.3.2. But while the traditional sliding window method moves gradually over the data set, in this thesis the window is jumped so that the step is the same as the window size, and so the windows don't overlap. This is illustrated below.

Normal sliding window method	Modified sliding window in this thesis
1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9
1 2 3 4 5 6 7 8 9	1 2 3 4 5 6 7 8 9

The two methods, **Longest All-Factors-Relationship** and **Two-Factors-Relationship** give different results. The **Longest All-Factors-Relationship** can tell what the relationships of all factors are, when one factor is focused upon at a point of time. While the **Two-Factors-Relationship** gives relationship information that exist between the main focused factor when it is

maximum or minimum, and another factor at a point of time (only two factors are compared at a point of time). In both cases, a factor is focused upon to find when this factor's maximum or minimum had occurred. Also, in both case other factor's changes are analyzed as positive (P), negative (N), or stable (S).

The discussions in chapter 6 are divided based on the sliding windows size, day or night data study, and minimum and maximum data study. Section 6.2 discusses the comparison of data from two sensors. Section 6.3 describes the software that was developed to analyze data based on a one-week window, using the frequent rules method. Section 6.4 discusses data analysis based on a 3-day window, with data taken 16 times a day. In sections 6.5 and 6.6, data analysis uses real hive data gathered in summer of 2018. Section 6.5 describes the two types of outputs and focuses on the Longest All-Factors-Relationship rules. Section 6.6 describes additional software written to include more periods and more sliding window sizes. Added studies include considering more data samples during a window, with 16 samples a day instead of only one point at noon in a 1-week window. Also, it involves adding a sliding window with 6 samples at night between 12 am to 2 am and focusing on different parameters when they are a minimum in the sliding window, instead of being maximum that was explored in previous sections of this chapter. Sections 6.3 to 6.6 focus on different parameters, while their amount of maximum or minimum is looked for in the sliding windows. Then maximum or minimum rules are made by finding the other parameters' relationship to this focused parameter. The real hive data, which is being analyzed [2] for this chapter and chapter 7 is from June 28, 2018 until August 22, 2018.

6.1. The structure of data analysis software program and generating mock data

The structure of the software program written for the data analysis employs a 3D array for data storage, since it helps to avoid using a lot of 2D Arrays or 1D arrays. Therefore, the datafile shown in section 4.3 is reformatted and stored into the structure of a 3D array. How this is done is discussed in section 6.1.2.

Also needed at the start of the software development, was a mock dataset to test the program since data form the hives was not yet available. The mock data set is discussed in section 6.1.1.

6.1.1. Mock data

A mock dataset was generated to test the data analysis programs. The mock dataset was generated for all sensors over a one-month period, with data created for every half an hour. The file consists of random numbers, and in some cases (for example for day to night temperature variation) a scheduled decrease and increase number is set is generated to help represent real data.

6.1.2. The structure of data analysis software

In this section, the analyzing software and how the data is stored in it is described. The original data file from the Arduino described in chapter 4.3 stored data in 2D arrays. The data is one-time complete reading of all sensors by the Arduino, taken every 30 minutes. This data is read and re-stored in a 3D array. By re-formatting and storing the data in this 3D array, there is no need for the data analysis software to repeatably read the data.

The structure of the 3D array is as follows. Consider **n** as the total number of complete sensor readings and having a 3D array [z] [x] [y]. Dimensions **x** and **y** are the data from one full reading of the sensors (see datafile of Figure 4.2) that would be stored in the Arduino, at a given time.

Dimension **x** (rows in Figure 4.2) is based on categories of the data (each sensor's data for example), and dimension **y** (column in Figure 4.2) is the on the stored data value. Dimension **z** is the time of the reading of the data stored in [x] and [y]. The below illustration shows one row of the **z** dimension, showing Z_0, the increasing **x** dimension index, and **y** dimension data values. The Z index increases for the next time the Arduino does a full reading.

Z_0:

    x_0: y_0, y_1, …,

    x_1: y_0, y_1. …,

    x_2: …,

    …

6.2. Comparing data from two sensors

In this section the software which does the comparison of data from two sensors is described. First, the user chooses which data variables to compare, and the software calculates the average or slope of the data. Consider the below description, and the flowchart in Figure 6.1.

Part 1) It requests that the user enters the file name. It tries to open the file.

Part 2) In case the file exists, it starts reading from it line by line. It stores the data in their related rows.

Part 3) It asks the user to enter their chosen functions like averages of two variables, a new period (start date and end date) other than first and last date, new name of file (for example the same format data file covering other data in other dates), start date and end date slope of two variables. The two variables can be between capacitance boards, its parts, temperatures or light sensor. Variables are capacitance boards 1 to 6, temperature sensors

1 and 2, and light sensor. The capacitance sensors can be interrogated based on the options of a whole capacitance board, right side, middle side or left side.

Part 4) If there is data in the interval period, the function that the user has chosen will be calculated for the two variables that the user has chosen.

Part 5) If the file does not exist, it will show an error message and ask for a correct name of the file. If the file exists and data reading is finished, it will show end of file data.

After answering the first request of the user, it will be asking again for average and slopes for variables along with some other functions like quitting or a new period (start date and end date) other than first and last date, new name of file (for example the same format data file covering other data in other dates).

In Appendix C, there is an example of the output of the two variable comparison software. The input of this software is the real hive data obtained from Valerie Beynon [2]. In Appendix D, there is another output example for comparing two temperature sensors by choosing a new period and file name.
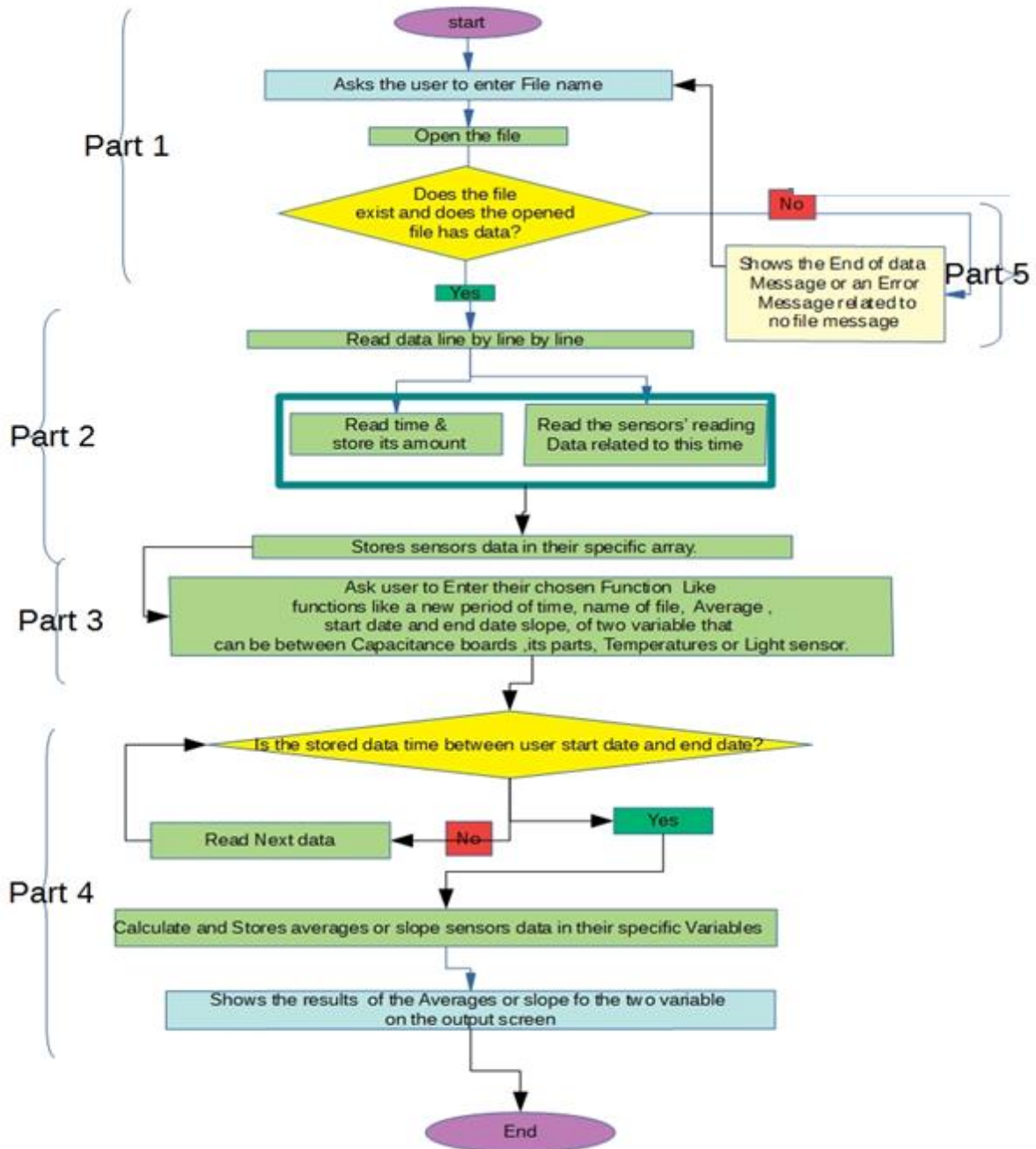
Figure 6.1- Flowchart for two parameter comparison software by Nadia Khoshgoo for output of Valerie Beynon 2018 research [2].

6.3. Software for 7-day-1-sample window data analysis

Because there is a lot of data obtained from the sensors every 30 minutes, one of the ways to explore and extract the relationship between factors can be extracting the **Longest All-Factors-Relationship** that may exist between the various sensors' data. A new combination of the frequent data mining and sliding window methods was used to extract the frequent rules between the data.

The software uses a sliding window data-mining technique to calculate the maximum data values taken at midday (12 pm) of a focused parameter, over a 1-week window. Then the difference to the day before is calculated.

One could ask why the average for the whole day was not used? Since using averages will include out-layers in the data, it was felt that this could spoil the data. As the goal is to count the frequencies of events, if extracting rules was based on average instead of maximum, outliers would affect the result and it would not produce actual rules. Therefore, by considering only maximum amounts, outliers will be omitted automatically after extracting rules, since the frequencies of outliers will be the lower.

One could also ask why maximum values were only considered. Maximization can be one the most important factors leading to something unexpected. Moreover, how bees react to a situation in which a variable is maximum can be interesting. Moreover, it is exciting to finding out what is the situation that is leading to their maximum activity in capacitance boards and what associated changes in other variables might have caused the maximum.

Consider the flowchart of Figure 6.2:

Part 1) It asks the user to enter the file name. It tries to open the file.

Part 2)   In case the file exists, it starts reading from it line by line. It stores the data in their related rows.

Part 3)   Light and capacitance sensors are focused on a way that their weekly window maximum amounts at noon are found. How much the value has changed from the previous day is calculated by amount and percentage. To find the factors that affect the focus factor's maximization, or how its maximization can have effect on the other factors' amount, the other factors percentage related to their previous amount is also calculated.

Part 4)   It stores the results obtained from part 3 in arrays assigned for the focused variable.

Part 5)   If the file does not exist, it will show the message error and ask for a correct name of the file. Also, it will show end of file data when the file reading has finished.
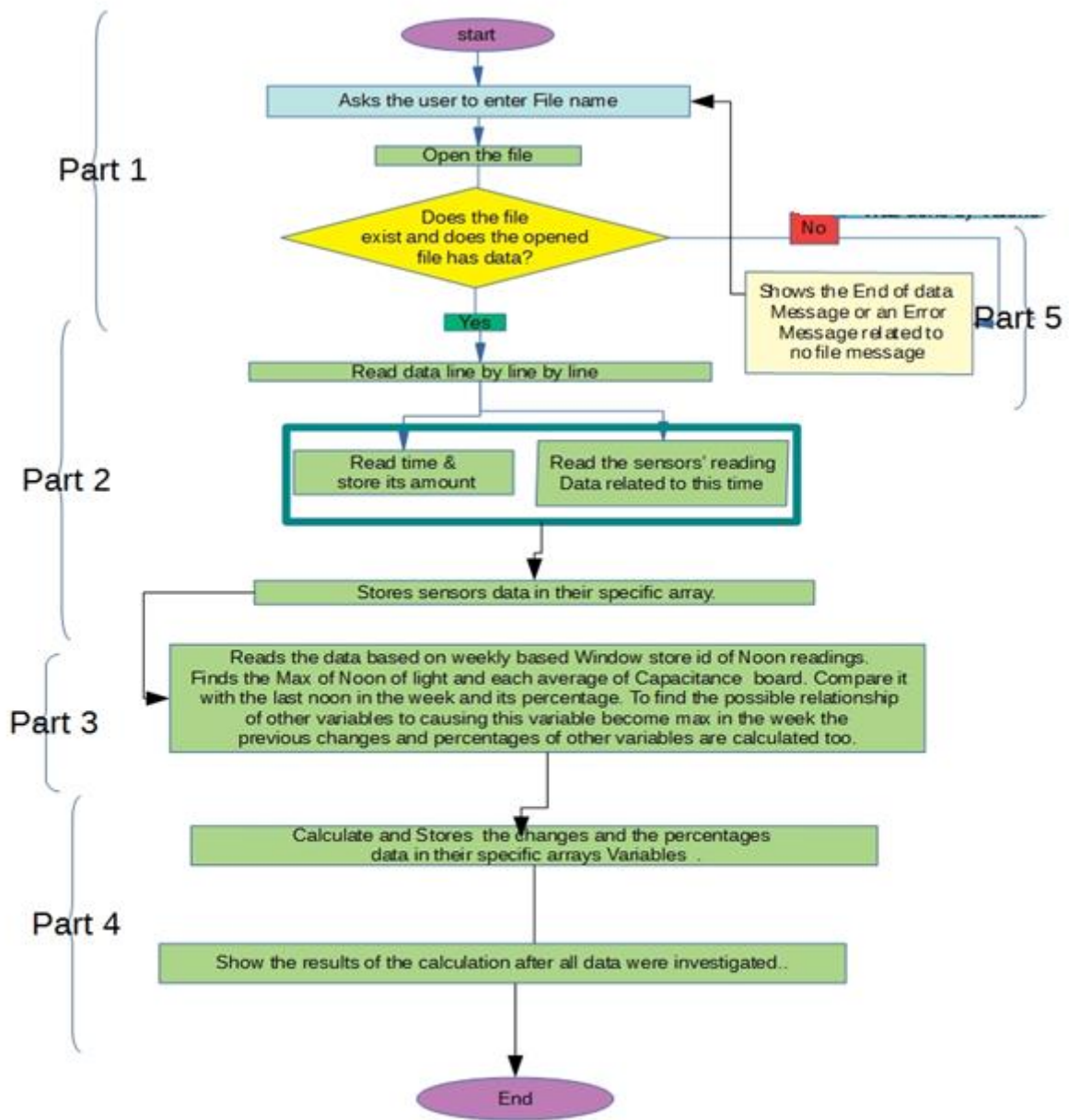
Figure 6.2- Flowchart for software of weekly sliding window frequent data analysis by Nadia Khoshgoo for output data of Valerie Beynon 2018 research [2].

6.4. Software for 3-day-16-sample window data analysis

Considering that 7-Day-1-Sample might not acquire enough data . For this reason, a 3-day data analysis was also considered. Also, to further increase the cases to be explored, data was evaluated 16 times per day (between 9 am and 4 pm) instead of 1 time per day.

The results of each rule of 16-time daily reading for maximum light and for maximum of each of the 6 capacitance boards were obtained. The all-rules-relationships of light-based maximum are compared to each other, and their occurrence frequencies are calculated. The extracted rules with frequencies will show what rules are the dominated rules and can show the **Longest All-Factors-Relationship** between the variables. The software program was tested both with real hive data, and mock data.

Consider the flow chart of Figure 6.3.

Part 1) It asks the user to enter the file name. It tries to open the file.

Part 2) In case it exists, it starts reading from it line by line. It stores the data in their related rows.

Part 3) Light and capacitance sensors are focused in a way that their 3-day window maximum at 16 samples a day (in the period of 9 am and 4 pm including) are found. How much it has changed from the previous day is calculated by amount and percentage. To find the affecting factors on its maximization or how its maximization can have effect on the other factors change when it is maximum, their percentage to their previous amount is calculated.

Part 4) It stores the results obtained from part 3 in specific arrays assigned for the focused variable.

Part 5)   If the file does not exist, it will show the message error and ask for a correct name of the

file. Also, it will show end of file data when the file reading has finished.
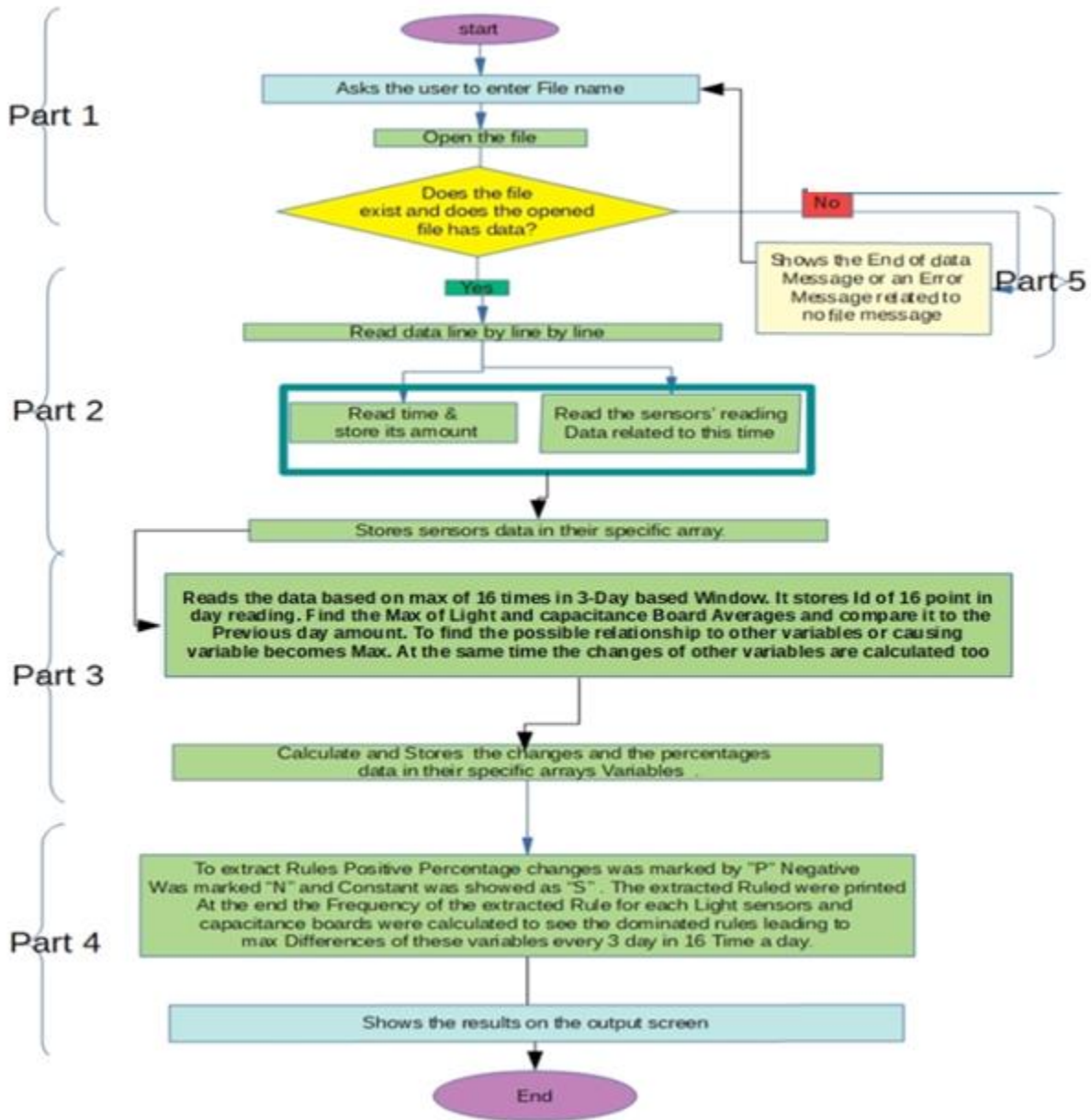
Figure 6.3- Flowchart for software of 3-day window data analysis and Forming All-relationship rules by Nadia Khoshgoo for data form Valerie Beynon 2018 research [2].

6.5.   Analysis of sliding window data

In section 6.5.1 the data output is analyzed based on a sliding window using the **Longest All-Factors-Relationship**. Section 6.5.2 explains how data is averaged or reformatted to make it suitable for analysis.

6.5.1.   Longest All-Factors-Relationship

All-Factors-Relationship's rules are rules that exist between the data gathered from various sensors. They are identified by how they change from the prior data amount. The sign for positive change is designated "P", for a constant one is "S", and for a negative one is "N". By attaching the signs gathered for all sensors together, I form a rule sentence. After analyzing the entire data set, the frequency that a rule may have happened is appended at the end of each rule sentences to show how common the rule is.

In the frequent itemset method commonly used in data mining, the items in the itemset are separate items. For example, they could be bought together in a transaction (this is described with an example in section 2.4). But in this thesis, each letter in the **Longest All-Factors-Relationship** rule sentence shows data changes in terms of sign instead of an item presence in a transaction (which would be the case in a frequent itemset). For example, in a system with 11 sensors, an all positive (P) changes for a **Longest All-Factors-Relationship** with a frequency of occurrence of 12 would be "PPPPPPPPPPP12". This example has 11 P's illustrating the 11 factors sign changes for this rule. In this **Longest All-Factors-Relationship**, each position in a rule sentence is only devoted to changes for one sensor and it is a fixed position. This is different from how itemset are considered in transactions in frequent itemset data mining, where any item can be in any position

in an itemset. This fixed position will help to compare the rules of various focused factors in the similar window size and similar sample number studies in sections 6.6 to 6.8.

It should be noted that in the sensor system of this thesis, since there are 11 sensors, the total probability of possible different rule sentences in the **Longest All-Factors-Relationship** is 3^11. However, it should be noted that in the case of the light sensor, the light measured at night can repetitively be 0, and so there would be no changes in state (S) for it at night. As a result, in the best case the maximum possible number of probable rule terms 2×3^10. Considering this large number of possible relationships, there is a need for obtaining the rules in software, since it is not possible to do this manually.

Illustration of the Longest All-Factors-Relationship implementation

The following example describes how the **Longest All-Factors-Relationship** is implemented. In this example data CaB refers to a capacitance board, Tem means temperature sensor, and Hum means humidity sensor.

For data taken at 9 am August 20, 2030. The data file would be:

```
Date: 20,08,2030
Light, CaB1, CaB2, CaB3, CaB4, CaB5, CaB6, Tem1, Hum1, Tem2, Hum2
500.00,22.10,23.20,22.23,20.21,22.10,23.20,30.00,40.00,32.00,43.00
```

For data taken at 9 am August 21, 2030. The data file would be:

```
Date: 21,08,2030
Light, CaB1, CaB2, CaB3, CaB4, CaB5, CaB6, Tem1, Hum1, Tem2, Hum2
502.00,23.10,24.20,21.23,22.21,23.10,21.20,31.00,42.00,33.00,49.00
```

For data taken at 9 am August 22, 2030. The data file would be:

```
Date: 22,08,2030
Light, CaB1, CaB2, CaB3, CaB4, CaB5, CaB6, Tem1, Hum1, Tem2, Hum2
480.00,21.10,22.20,22.40,20.21,22.10,21.20,31.00,43.00,33.00,49.00
```

By focusing on the Light as the focus factor, it is maximum on August 21, 2030. The rule sentence that will be obtained considering the differential change from the proceeding day (August 20) is determined as follows:

Light change: 502.00 - 500.00 = 2 > 0 → the sign of the change is **P** (positive), so the rule is **P**

Capacitance board 1 change: 23.10 - 22.10 = 1 > 0 → the sign of the change **P**, so the rule is **P**

Capacitance board 2 change: 24.20 - 23.20 = 2 > 0 → the sign of the change **P**, so the rule is **P**

Capacitance board 3 change: 21.23 - 22.23 = -1 <0 → the sign of the change **N**, so the rule is **N**

Capacitance board 4 change: 22.21 - 20.21 = 2 > 0 → the sign of the change **P**, so the rule is **P**

Capacitance board 5 change: 22.10 - 23.10 = 1 > 0 → the sign of the change **P**, so the rule is **P**

Capacitance board 6 change: 21.20 - 23.20 = -2 < 0 → the sign of the change **N**, so the rule is **N**

Temperature 1 change: 30.00 - 31.00 = 1 > 0 → the sign of the change **P**, so the rule is **P**

Humidity 1 change: 42.00 - 41.00 = 1 > 0 → the sign of the change **P**, so the rule is **P**

Temperature 2 change: 33.00 - 32.00 = 1 > 0 → the sign of the change **P**, so the rule is **P**

Humidity 2 change: 49.00 - 43.00 = 6 > 0 → the sign of the change **P**, so the rule is **P**

The rule sentence is then formed by attaching all the sign changes together, which forms the rule **PPPNPPNPPPP**. Since only one comparison has been looked at, the frequency is 1, and so the frequency of the rule sentence is 1. If for example, much more data would be explored and the same rule sentence occurred again, the frequency number would increase appropriately.

Another illustration is shown in the below table, but in this case, the focus is on the maximum situation for capacitance board 3. Here, the data changes result in the rule sentence for capacitance board 3 to be **NNNPNNSSPSS,** and as it has only happened once it is **NNNPNNSSPSS1.**

Table 6.1. Longest All-Factors-Relationship frequency calculation table with focus on maximum of capacitance board 3. The humidity senor has +/- 3 relative humidity with range 0-80% RH and the temperature sensor has +/-0.4 C temperature accuracy with range -10 to +85 C. [30]

Sensor / Date	Light (lux)	CaB1 (pF)	CaB2 (pF)	CaB3 (pF)	CaB4 (pF)	CaB5 (pF)	CaB6 (pF)	Tem1 (C)	Hum1 (RH)	Tem2 (C)	Hum2 (RH)
20,08,2030	500.00	22.10	23.20	22.23	20.21	22.10	23.20	30.00	40.00	32.00	43.00
21,08,2030	502.00	23.10	24.20	21.23	22.21	23.10	21.20	31.00	42.00	33.00	49.00
22,08,2030	480.00	21.10	22.20	22.40	20.21	22.10	21.20	31.00	43.00	33.00	49.00
Data change from 22nd -21^{at}	480 - 502 = -22<0	21.10 - 23.10 = -2<0	22.20 - 24,20 = -2<0	22.40 - 21.23 = 1.17>0	20.21 - 22.21 = -2<0	22.10 - 23.10 = -1<0	21.20 - 21.20 = 0	31 - 31 = 0	43 - 42 = 1>0	33 - 33 = 0	49 - 49 = 0
Sign of data changes	N	N	N	P	N	N	S	S	P	S	S

Application of the Longest All-Factors-Relationship to data obtained from the hive:

The below rule sentences were generated from the 7-day-1-sample window software analysis of the real Hive data, for the situation where capacitance board 1 is the focus.

```
Wholecapacitance 1: NPNNNNNNNNN1
Wholecapacitance 1: NPPPPPPPNPP1
Wholecapacitance 1: PPNPNNPPPNN1
Wholecapacitance 1: NPPNPPNPNNN1
Wholecapacitance 1: NPPPPPNPNNN1
Wholecapacitance 1: SPPPPPNNPPP1
Wholecapacitance 1: NPPPPPPPPNN1
```

Looking at these rule sentences, it can be seen that the frequency of events is 1. This low frequency of occurrence was common of the 7-day-1-sample data analysis, and so it was concluded that more data samples were needed. This was done by employing the 3-day-16-sample window, with 16 samples a day.

6.5.2. Evaluation and discussion of sensor data

In this section, to evaluate the sensors' data described in chapter 4, the input file is explored and reformatted to become ready and suitable to be exported to Excel.

Preparing the previously formatted input to be exported to Excel

The real hive data described in chapter 4 includes all the sensors' outputs in a way that each sensor data is in a separate row and in common columns. Therefore, the data file was reformatted to enable importation to Excel, such that in each row of data there is the data from a one time reading from all sensors, and each column has the data of each individual sensor or capacitance board. In this format, the value shown for each capacitance board is the average of its 12 capacitance sensors. This file format is shown in Figure 6.4.

Figure 27 plots the data that which is shown in Figure 6.4, for the capacitance boards. It can be seen that the capacitance averages for each capacitance board (labeled as CB followed by their capacitance board number such as CB1, CB2, etc.) have daily fluctuations. However, clearly capacitance boards 2 and 5 show an overall increasing trend to their data. Capacitance board 5 shows increasing capacitance signal between July 23 to August 22, 2018, and capacitance board 2 shows increasing signal between the July 10 to August 1, 2018. These were the capacitance boards upon which the bees grew honeycomb (board 5 was in the honey super, and board 2 was in the brood box).

From inspection of the data, however, It can be concluded that finding clear correlation relationships between events occurring between the sensors and capacitance boards is not easy. Therefore, in order to determine possible correlations, various data analysis methods were

performed on the original hive data described in chapter 4. This will be explored in following section.
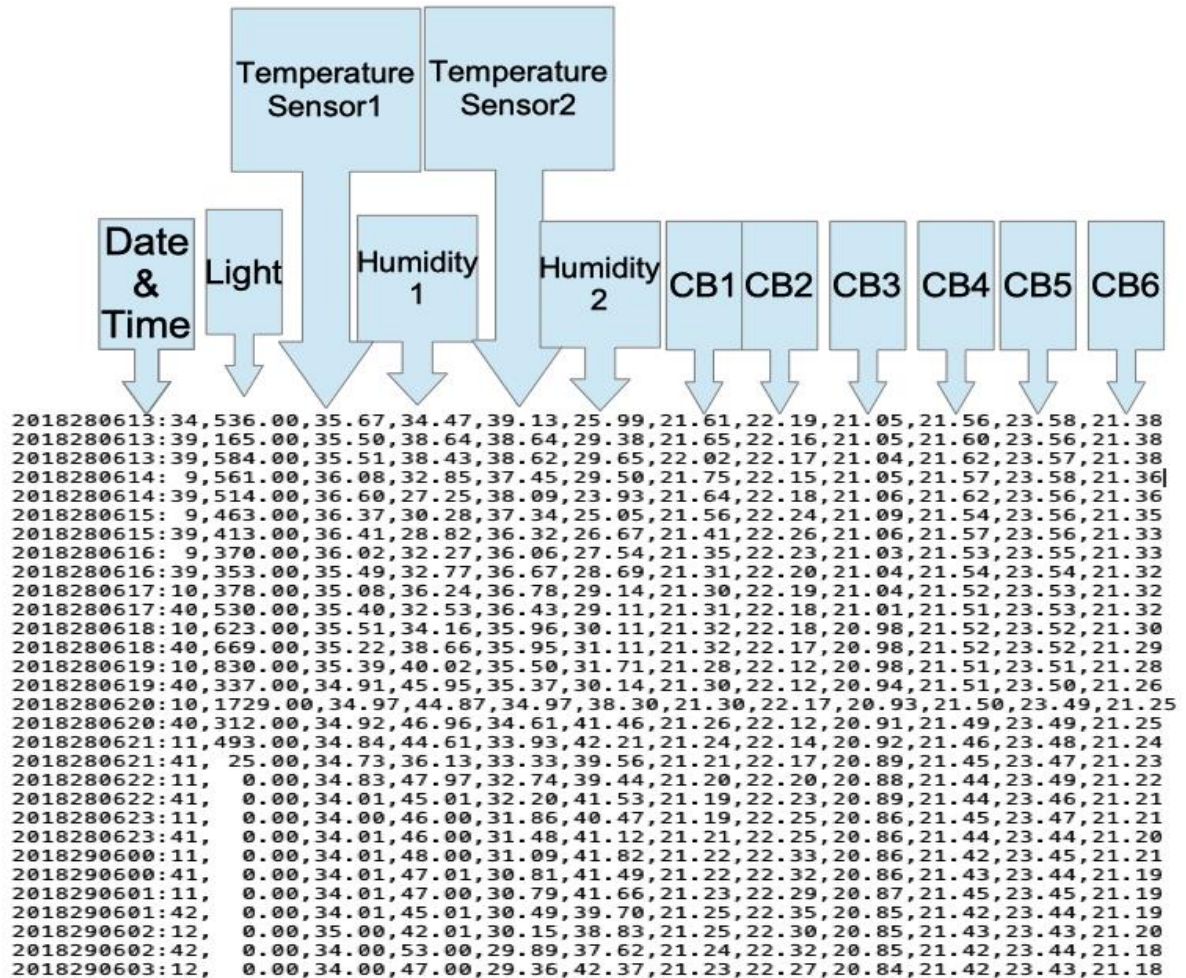
Figure 6.4- New formatted and edited input for drawing linear charts. The software that produces this file uses the file described in chapter 4 as its input.
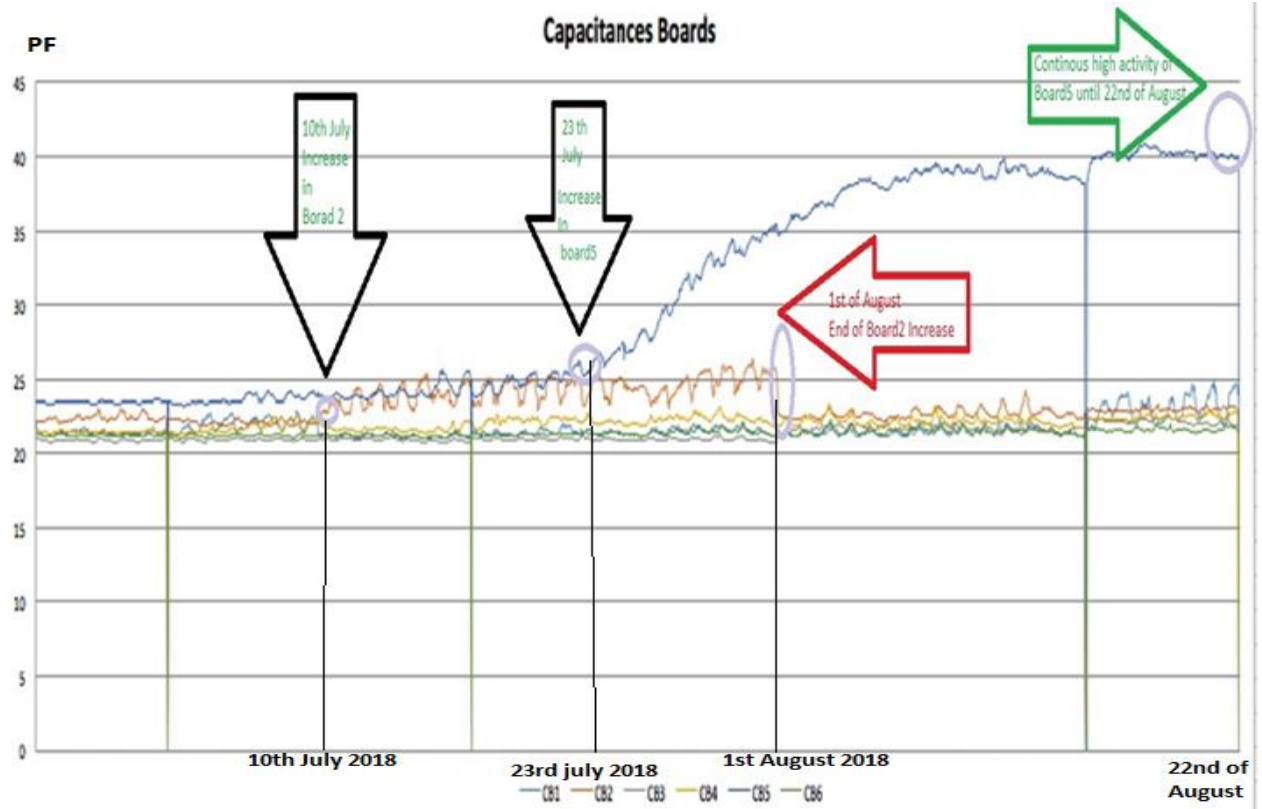
Figure 6.5- Plot of capacitance boards' data average sample of data is shown in Figure 6.4.

6.6. Results of the data mining done for phase 2 of the project

This section discusses the results from the hive data done employing the **All-Factors-Relationship's rules**. Two tables are presented to cover the dominant rules found, and these tables summarize the detailed bar charts, which are shown in Appendix E. The program covered both maximum and minimum rules in night or day in various window size and sampling number of periods. The program considers only the changes that are minimum and decreasing, or maximum and increasing, when compared to previous day, which maybe a last day in the previous window.

6.6.1. Rules equal and larger than a minimum frequency

In order to omit outliers in the data, a minimum criterion is applied to the rules found by the **All-Factors-Relationship's rules**. In this way, the rules that have the frequency more or equal than a defined minimum frequency are kept. These discovered rules have been drawn in bar charts and are available and described in Appendix E. Below is a sample of the All-Factors-Relationship rules found by the software for cases where minimum frequency is >=4.

```
Wholecapacitance 1, NPPNPPNPPPP4
Wholecapacitance 1, NPPPPPPPPPP6
Wholecapacitance 1, NPPPPPPPNPP4
Wholecapacitance 1, PPNNNNNNNPP7
Wholecapacitance 1, PPNNPPNNPPP4
Wholecapacitance 1, PPPNPPPPPPP5
Wholecapacitance 1, NPPNPPPPPPP6
Wholecapacitance 1, PPPPNPPNNPP5
Wholecapacitance 1, NPNNNPNNNPP5
Wholecapacitance 1, NPPNNPNNNPP5
Wholecapacitance 1, NPNPNNNNNNN5
Wholecapacitance 1, PPPPNPNPPNN4
Wholecapacitance 1, NPPNPPPPNPP4
Wholecapacitance 1, PPPPNNPPPNN4
```

```
Wholecapacitance 1, PPNNNNNNNNNN4
```

6.6.2. All-Factors-Relationship's rules charts results

Figure 6.6 illustrates the maximum activity rules during the day when the window size is 3-day-16-sample. It was observed that of the 6 capacitance boards, most changes were in capacitance boards 2 and 5. Table 6.2 illustrates their dominant relationship rules, showing their higher maximum activity (CaBn where 'n' are the capacitance boards' numbers, Tem1=temperature1 Hum1=humidity1 Tem2=temperature2 Hum2=humidity2).

Table 6.2. 3-day-16-sample maximum activity rules of capacitance boards 2 and 5.

Rule ＼ Sensors	Light	CaB1	CaB2	CaB3	CaB4	CaB5	CaB6	Tem1	Hum1	Tem2	Hum2	Frequency	Final Dominant rule
Dominant rule of capacitance board 2	P	P	P	P	P	P	P	P	P	P	P	7	PPPPPPPPPPP7
Dominant rule of capacitance board 5	P	P	N	P	P	P	P	N	P	P	P	9	PPNPPPPNPPP9
Dominant rule of capacitance board 5	N	P	P	P	P	P	P	P	P	P	P	7	NPPPPPPPPPP7
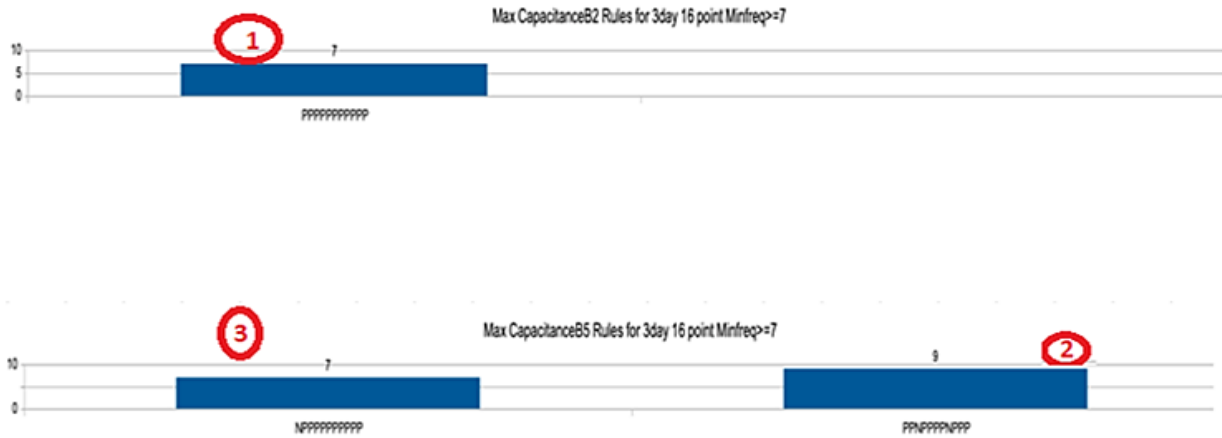
Figure 6.6- Maximum capacitance activity rules of capacitance boards 2 and 5.

The following tables 6.3 to 6.6 are the summary of the charts rules which are in Appendix E. In these tables, the common rules that are found are shown. Above each rule, the expression L123456THthF and L123456THth are shown, to help illustrate which sensor give that sign (P/N/S). The letters of L123456THthF represent each sensor (L=light, 1-6= are the capacitance board's numbers, T=Temmperature1 H=Humidity1, t=temperature2 h=humidity2), and the F is the frequency.

Dominant maximum All-Factors-Relationship rules, both day and night tables

Table 6.3 explores the maximum day rules. In exploring the results shown in Table 6.3. The following points can be seen .

- Conclusion 1: comparing the columns having rule Category ID (RCID) R1 and R2, it can be seen that there are more common maximizing rules between various sensors for the 3-day-16-sample case, compared to the 7-day-16-sample case.

- Conclusion 2: considering the column RCID R3, it can be seen that a common rule that occurs for both 3-day and 7-day sampling is the rule PPPPPPPPPPP. This occurs when

capacitance board 2 maximum-rules are the focus, showing that all sensor changes are also positive.

- Conclusion 3: considering the column RCID R3, it can be seen that a common rule that occurs for both 3-day and 7-day sampling is rule PPNPPPPNPPP. This occurs when capacitance board 5 maximum-rules is the focus, and capacitance board 2 activity and temperature 1 are decreasing and other factors are increasing.

Table 6.4 explores the maximum night rules. In exploring the results shown in Table the following points can be seen .

- Conclusion 4: comparing the columns having RCID R4 and R5, it can be seen that there are more common maximizing rules between various sensors for the 3-night-6-sample case, compared to the 7-night-6-sample case.

Table 6.3. Sensors' maximum dominant day rules and their occurrences for other sensors or in other sliding windows sizes. L123456THthF and L123456THth are their below rule's element title.

Rule / Sensor Dominant rule+ Frequency	7-day-16-samples rules	3-day-16-samples rules	Common dominant rules in maximum 3-day and 7-day
Dominant in Light	L123456THthF PPPPPPPPPP5	L123456THthF PPNPPPPNPPP12	
and also, in	CaB2	CaB4 and CaB5	
Dominant in CaB1	L123456THthF PPNNNNNNNNN5	L123456THthF PPNNNNNNNPP7	
Dominant CaB2	L123456THthF PPPPPPPPPPP3	L123456THthF PPPPPPPPPPP7	L123456THth PPPPPPPPPPP
and also, in	Light	CaB3, CaB4, CaB6	
Another dominant CaB2 rule is	L123456THthF NPPPNNPPPPP3		
Dominant in CaB3	L123456THthF SSSSSSSSSSS8	L123456THthF PPPPPPPPPPP11	
and also, in		CaB2, CaB4, CaB6	
Another dominant CaB3 rule is	L123456THthF PPPPPPPPPNN7		
Dominant in CaB4	L123456THthF PPPPPPPPPNP4	L123456THthF PPPPPPPPPPP7	
and also, in		CaB2, CaB3, CaB6	
Another dominant CaB4 rule is	L123456THthF NPPNPPPPPPN4	L123456THthF PPNPPPPNPPP7	
and also, in		Light, CaB5	
Another dominant CaB4 rule is	L123456THthF PPPPPPPPNPN4		
Dominant in CaB5	L123456THthF PPNPPPPPPPP5	L123456THthF PPNPPPPNPPP9	L123456THth PPNPPPPNPPP
Another dominant CaB5 is	L123456THthF PPNPPPPPPNP4	L123456THthF NPPPPPPPPPP7	
and also, in		Light, CaB4	
Another dominant CaB5 is	L123456THthF PPNPPPPNPPP4		
Dominant in CaB6	L123456THthF PPPNPPPPPPP7	L123456THthF NPPNPPPPPPP9	
Another dominant CaB6 rule is		L123456THthF PPPNPPPPPPN9	
Another dominant CaB6 rule is		L123456THthF PPPPPPPPPPP7	
and also, in		CaB2, CaB3, CaB4	
Rules Category ID	R1	R2	R3

Table 6.4. Sensors' maximum dominant night rules and their occurrences for other sensors or in other sliding windows sizes. L123456THthF and L123456THth are their below rule's element title.

Rule Sensor Dominant rule+ Frequency	7-night-16-samples rules	3-night-16-samples rules	Common dominant rules in maximum 3-night and 7-night
Dominant in Light and also in	L123456THthF SPPPPNPPPPP7	L123456THthF SPPPPPPPPPP12 CaB1, CaB2, CaB3, CaB4, CaB5, CaB6	
Dominant in CaB1 and also in Another dominant CaB1 rule is and also in	L123456THthF SPPPPPPPPPP8 CaB2, CaB4, CaB5, CaB6	L123456THthF SPPPPPPPPPP21 Light, CaB2, CaB3, CaB4, CaB5, CaB6 L123456THthF SPPPPNPPPPP10 CaB4	L123456THth SPPPPPPPPPP
Dominant in CaB2 and also, in	L123456THthF SPPPPPPPPPP4 CaB2, CaB4, CaB5, CaB6	L123456THthF SPPPPPPPPPP10 Light, CaB1, CaB3, CaB4, CaB5, CaB6	L123456THth SPPPPPPPPPP
Dominant in CaB3 and also in	L123456THthF SPPPPPPPPNP5	L123456THthF SPPPPPPPPPP18 Light, CaB1, CaB2, CaB4, CaB5, CaB6	
Dominant in CaB4 and also in Another dominant CaB4 rule is and also, in	L123456THthF SPPPPPPPPPP11 CaB1, CaB2, CaB5, CaB6	L123456THthF SPPPPPPPPPP27 Light, CaB1, CaB2, CaB3, CaB5, CaB6 L123456THthF SPPPPNPPPPP10 CaB1	L123456THth SPPPPPPPPPP
Dominant in CaB5 and also in Another dominant CaB5 rule is	L123456THthF SPPPPPPPPPP4 CaB1, CaB2, CaB4, CaB6 L123456THthF SPNPPPPNPPP6	L123456THthF SPPPPPPPPPP15 Light, CaB1, CaB2, CaB3, CaB4, CaB6	L123456THth SPPPPPPPPPP
Dominant in CaB6 and also in	L123456THthF SPPPPPPPPPP7 CaB1, CaB2, CaB4, CaB5,	L123456THthF SPPPPPPPPPP22 Light, CaB1, CaB2, CaB3, CaB4, CaB5	L123456THth SPPPPPPPPPP
rules Category ID	**R4**	**R5**	**R6**

Dominant minimum All-Factors-Relationship rules, both day and night Tables

Table 6.5 explores the minimum day rules. In exploring the results shown in Table 6.5 the following points can be seen.

- Conclusion 1: for the daytime minimum rules, comparing the columns having rule Category ID R7 and R8, when the window size is reduced from 7-day to 3-day, the number of common rules has decreased, which in its turn can show that each capacitance board activity is minimized based on a specific rule in a day.

- Conclusion 2: considering the column RCID R9, it can be seen that a common rule that occurs for both 3-day and 7-day sampling is the rule NNPNNPNNNPP, which occurs when the light and capacitance board 3 are the focus. This occurs when CaB2 and CaB5 and Temperature 2 and Humidity 2 have increased.

Table 6.6 explores the minimum night rules. In exploring the results shown in Table 6.6 the following can be seen.

- Conclusion 3: for the nighttime columns having rule Category ID R10 and R11, it can be seen that there are more common minimizing rules between various sensors when the window size has been 3-night instead of 7-night.

- Conclusion 4: for the column RCID R12, the common dominant rule for CaB1, CaB2, CaB3, CaB4, CaB5, CaB6 for both 3-day and 7-day is SNNNNNNNNNN. This dominant minimum activity rule for capacitance boards happens when the other capacitance board activities have decreased.

Table 6.5. Sensors' minimum dominant day rules and their occurrences for other sensors or in other sliding windows sizes. L123456THthF and L123456THth are their below rule's element title.

Rule / Sensor Dominant rule+ Frequency	7-day-16-samples rules	3-day-16-samples rules	Common dominant rules in minimum 3-day and 7-day
Dominant in Light **and also in**	L123456THthF NNPNNPNNNPP7 CaB3	L123456THthF NNPNNPNNNPP9 CaB3	L123456THth NNPNNPNNNPP
Dominant in CaB1 **and also in**	L123456THthF PNNPNPNNNNN9 CaB4, CaB6	L123456THthF NNPNNPNNNNN11 CaB4	
Dominant in CaB2 **Another dominant CaB2 rule is** **Another dominant CaB2 rule is**	L123456THthF PNNNNNNNNNN3 L123456THthF SSSSSSSSSSS3 L123456THthF NNNPNNPNPNP3	L123456THthF NNNPPPPPPNP6 L123456THthF PNNPPPPPPNP7 L123456THthF SSSSSSSSSSS9	
Dominant in CaB3 **and also, in**	L123456THthF NNPNNPNNNPP8 Light	L123456THthF NNPNNPNNNPP10 light	L123456THth NNPNNPNNNPP
Dominant in CaB4 **and also in**	L123456THthF PNNPNPNNNNN8 CaB1, CaB6	L123456THthF NNPNNPNNNNN11 CaB1	
Dominant in CaB5 **Another dominant CaB5 rule is**	L123456THthF PNNPNNNNNNP3 L123456THthF NNNPNNNNNNP3	L123456THthF PNNPNNNNNNN6	
Dominant in CaB1 **and also in**	L123456THthF PNNPNPNNNNN9 CaB1, CaB4	L123456THthF PPPNPPPPPPP9	
rules Category ID	**R7**	**R8**	**R9**

Table 6.6. Sensors' minimum dominant night rules and their occurrences for other sensors or in other sliding windows sizes. L123456THthF and L123456THth are their below rule's element title.

Rule / Sensor Dominant rule+ Frequency	7-night-6-samples	3-night-6-samples	Common dominant rules in minimum 3-night and 7-night
Dominant in Light	L123456THthF SPPPPNPPPPP7	L123456THthF SNNNNNNNNNN10	
also dominant in		CaB1, CaB2, CaB3, CaB4, CaB5, CaB6	
Another dominant rule is		L123456THthF SNPNNNNNNNN10	
also dominant in		CaB3, CaB5, CaB6	
Dominant in CaB1	L123456THthF SNNNNNNNNNN6	L123456THthF SNNNNNNNNNN10	L123456THth SNNNNNNNNNN
also dominant in	CaB2, CaB3, CaB4, CaB5, CaB6	Light, CaB2, CaB3, CaB4, CaB5, CaB6	
Dominant in CaB2	L123456THthF SNNNNNNNNNN9	L123456THthF SNNNNNNNNNN11	L123456THth SNNNNNNNNNN
also dominant in	CaB1, CaB3, CaB4,CaB5, CaB6	Light, CaB1, CaB3, CaB4, CaB5, CaB6	
Another dominant rule is	L123456THthF SNNPNPPPPNP4		
Dominant in CaB3	L123456THthF SNNNNNNNNNN4	L123456THthF SNNNNNNNNNN11	L123456THth SNNNNNNNNNN
also dominant in	CaB1, CaB2, CaB3, CaB5, CaB6	Light, CaB1, CaB2, CaB3, CaB5, CaB6	
Another dominant rule is	L123456THthF SNNNNPNNNPN5	L123456THthF SNPNNNNNNNN8	
also dominant in		Light, CaB5, CaB6	
Dominant in CaB4	L123456THthF SNNNNNNNNNN9	L123456THthF SNNNNNNNNNN11	L123456THth SNNNNNNNNNN
also dominant in	CaB1, CaB2, CaB3, CaB5, CaB6	Light, CaB1, CaB2, CaB3, CaB5, CaB6CaB1	
Dominant in CaB5	L123456THthF SNNNNNNNNNN4	L123456THthF SNNNNNNNNNN9	L123456THth SNNNNNNNNNN
also dominant in	CaB1, CaB2, CaB3, CaB4, CaB6	Light, CaB1, CaB2, CaB3, CaB4, CaB6	
Another dominant rule is	L123456THthF SNPNNNPNNN6	L123456THthF SNPNNNNNNNN6	
also dominant in		Light, CaB3, CaB6	
Dominant in CaB6	L123456THthF SNNNNNNNNNN5	L123456THthF SNNNNNNNNNN11	L123456THth SNNNNNNNNNN
also dominant in	CaB1, CaB2, CaB3, CaB4, CaB5	Light, CaB1, CaB2, CaB3, CaB4, CaB5	
Another dominant CaB6 rule is		L123456THthF SNPNNNNNNNN6	
also dominant in		Light, CaB3, CaB5	
rules Category ID	R10	R11	R12

Comparing minimum rules for 3-night with 7-night Window

By comparing the minimum rules for 3-night and 7-night windows in the Appendix E charts, it became apparent that reducing the size of sliding window (from 7-day to 3-day) leads to finding a common dominant global rule being revealed, which is in this case SNNNNNNNNNN. This rule became dominant and more apparent in 3-night in (Appendix E) for the capacitance board 3 and capacitance board 5 than in the 7-night. However, in 7-night (Appendix E) this rule was the second dominant rule.

6.7.   In depth comparison of 3-day vs 7-day analysis sliding windows

When examining the charts in Appendix E, it can be seen that the 7-day-1-sample analysis generally did not have enough samples for extracting generalizing rules. By comparing the maximum day analysis using the 3-day-16-sample and the 7-day-16-sample, it can be seen that having the 16 samples per day gave more times in a day when the data was analyzed, and this gave beneficial results.

Since in the sampling every 3-days resulted in more cases reviewed, the 3-day-16-sample windows made the relationship more apparent and more suitable for finding the dominant common rules. However, a few specific rules could still be obtained from 7-day-16-sample windows.

Considering minimum activity situations for both 7-day and 3-day sampling, it was found that they both give good results. For example, both show capacitance board 1 and capacitance board 4 activity having a minimum with temperature drop.

Comparing maximum night rules, 3-night window shows better main global dominant rule, which is SPPPPPPPPPPP, which has become apparent. Also, the increase in capacitance board 1 and capacitance board 4 activity with decrease in capacitance board 5 in "SPPPPNPPPPP".

However, other rules with close frequency amounts to each other can be gained from 7-night rules. Considering, minimum night rule when the window sized was reduced from 7-night to 3-night, it is possible to see a global common rule of "SNNNNNNNNNNN", showing that 3-night works better.

Generally, for all periods if a rule is dominant, more exploring rules can be achieved by reducing the size of the sliding window. This can reveal whether a rule is global or not.

Table 6.7 gives a summary of results for the methods with various windows sizes and periods. In Table 6.7 it can be seen that the night the 3-night-6-sample method gave common global rules for both maximum-rules and minimum rules. Therefore, decreasing the window size helps in getting a global frequency rule. Also, reducing the window size from 7-day-16-samples to 3-day-16-samples for daytime data, helps get several common frequency rules for some capacitance boards.

Table 6.7. Summary of data analysis methods obtained for getting All-Factor frequent rules.

Obtained rules / Window size and samples	Global Longest rule	Common Dominant The longest rules between capacitance boards	Common Second Dominant The longest rules between capacitance boards	Common First and Second Dominant The longest rules between capacitance boards	Specific Dominant Longest rules of boards (not common)
maximum 7-day-1-sample	No (not enough samples were reviewed)	No (not enough samples were reviewed)	No (not enough samples were reviewed)	No (not enough samples were reviewed)	No (not enough samples were reviewed)
maximum 7-day-16-sample	-	-	-	-	Yes
maximum 3-day-16-sample	-	Yes, for several cases	Yes	Yes	Yes
minimum 7-day-16-sample	-	Yes, for few cases	It may exist, but it is not apparent	It may exist, but it is not apparent	Yes
minimum 3-day-16-sample	-	Yes, for several cases	Yes	Yes	Yes
maximum 7-night-6-sample	-	Yes, for several cases	Yes	Yes	Yes
maximum 3-night-6-sample	Yes	Yes	Yes	-	-
minimum 7-night-6-sample	-	Yes, for several cases	Yes	Yes	Yes
minimum 3-night-6-sample	Yes	Yes, for several cases	Yes	Yes	Yes

6.8.   Summary of Linear chart vs All-Factors data analysis results

In this chapter it can be seen that the overall trend of capacitance boards and other sensors fluctuations through the linear chart drawn from the edited input file. However, by looking at these charts, the calculated relationship and rules between all elements together is difficult and challenging to extract. And so, the data analysis routines for extracting rules between them was developed, which gave various output results that show the rules existing between sensor data in different periods of time. It could be helpful for understanding the relationship between various factors, if repeated rules would be explored through time in the sliding windows. Considering both maximum and minimum rules helps to see what happens at a point of time spanning two days, when a factor becomes maximum or minimum. Considering all factors should be more reliable, as some factors may together cause a situation, or a situation may cause some factors to change. In cases for maximum condition, a 3-day window gave fewer frequent rules, which helps to see the trends better. Some rules that were extracted for maximum day rules for 3-day-16-sample were like, PPPPPPPPPP for maximum capacitance board 2, 4 and 3. Moreover, PPNPPPPNPPP was for maximum light. It may indicate Light maximization may lead to capacitance board 2 (brood board) activity decrease. Also, it was one of the maximum rules of capacitance boards 4 and 5, PPNPPPPNPPP, may show reduction in temperature1 and capacitance board 2 activity amount caused capacitance board 4 and 5 activity maximization.

Considering minimum rules for 3-day-16-point, NNPNNPNNNPP was for capacitance board 3 and NNPNNPNNNNN for capacitance boards 1 and 4. Considering these two rules, it seems an increase in temperature2 has decreased activity in capacitance board 3 (NNPNNPNNNPP), and caused minimum activity in capacitance board 3. A decrease in temperature sensors (NNPNNPNNNNN) decreased activity in capacitance board 1 and 4 and caused minimum activity

92

in capacitance board 1 and 4. This may show that capacitance board 1 and 4 are mostly used as a shelter place in high environment temperature. It may be because they were on the left side of the hive, which had trees shading them from the sun, and so making them a shelter place from heat. For minimum night rules SNNNNNNNNNN was the dominant one in 3-night 6-samples for all capacitance sensors.

Chapter 7. Phase 2: Two-Factors-Relationship software outputs chart PN (Positive Negative) bar charts

7.1.   Two-Factors-Relationship definition

In the previous chapter, the relationships which are based on the sensor data changes, and relationship rules were discovered based on different analysis methods, was shown. The method used in this chapter is to compare how one factor changes with respect to other factors, one by one (not in a group and not all factors like the All-Factors-Relationships in chapter 6). This method is named Two-Factors-Relationship.

In the presented Two-Factors-Relationship, the changes of a focused factor with respect to another factor, and their relationship's frequencies, is found. This information is shown in a two-color bar-chart that is called a PN bar-charts (Positive and Negative changes bar-charts). The Two-Factors-Relationship considers two variables, and data change between two values are given a positive or negative sign, which is resulted from calculating the difference between the variable at two different days. The changes that are $>= 0$ are designated P, and changes that are< 0 are designated N.

An example is shown in Table 7.1 for some mock data for a 3-day sliding window (from August 20, 2030 to August 22, 2030), for capacitance board data with the 12 sensors averaged into one value. In Table 7.1, the mock data CaB means capacitance board, Te means Temperature sensor, and Hu means Humidity sensor. Focusing on the data for capacitance board 3, showed that it is maximum on the 22nd of August 2030. The comparative value (**P** or **N**) of all sensors is then between August 21 and August 22. The method counts the frequencies of **P** or **N** occurrences, and so the show the incremental calculation of N+1 or P+1 as appropriate.

Table 7.1. Two-Factors-Relationship frequency calculation table for maximum of capacitance board 3 when it is the focus. It is using a mock data made up for a 9 o'clock sensor reading.

Sensor Date	Light (lux)	CaB1 (pF)	CaB2 (pF)	CaB3 (pF)	CaB4 (pF)	CaB5 (pF)	CaB6 (pF)	Te1 (C)	Hu1 (RH)	Te2 (C)	Hu2 (RH)
20,08,2030	500.0	22.10	23.20	22.23	20.21	22.10	23.20	30.00	40.00	32.00	43.00
21,08,2030	502.0	23.10	24.20	21.23	22.21	23.10	21.20	31.00	42.00	33.00	49.00
22,08,2030	480.0	21.10	22.20	22.40	20.21	22.10	21.20	31.00	43.00	33.00	49.00
Data Changes of 22nd-21nd	-22	-2	-2	1.17	-2	-1	0	0	1	0	0
P and N frequencies changes	N+1	N+1	N+1	P+1	N+1	N+1	P+1	P+1	P+1	P+1	P+1

## 7.2. Positive-Negative changes (PN) frequency bar chart

In this section the positive negative changes (PN) frequency bar charts are shown for when light is the focus. The frequencies of P or N changes of other factors is calculated. The various charts show the data for different cases, with light as the maximum, light as the minimum, and for different time windows.

Considering that there would be usually a change between two days of data for another factor than light, the zero changes are usually not happening unless the sensor connection is lost.

7.2.1. PN charts focusing on maximum light at daytime

Considering figures 7.1, 7.2, and 7.3, in almost all capacitance boards an increase in activity when the light is maximum could be seen. In only a few cases (points) for the 7-day-1-sample of Figure 7.1, a decrease in activity in capacitance board 2 could be seen. However, this could not be seen for the larger time window of figures 7.2 and 7.3. Therefore, the negative result of the 7-day-1-sample about capacitance board 2 can be said to be not definite. In all these figures 7.1, 7.2 and

7.3, it is shown that the largest differential between increase and decrease frequency counts are in capacitance board 1, capacitance board 3, capacitance board 5, and capacitance board 6.
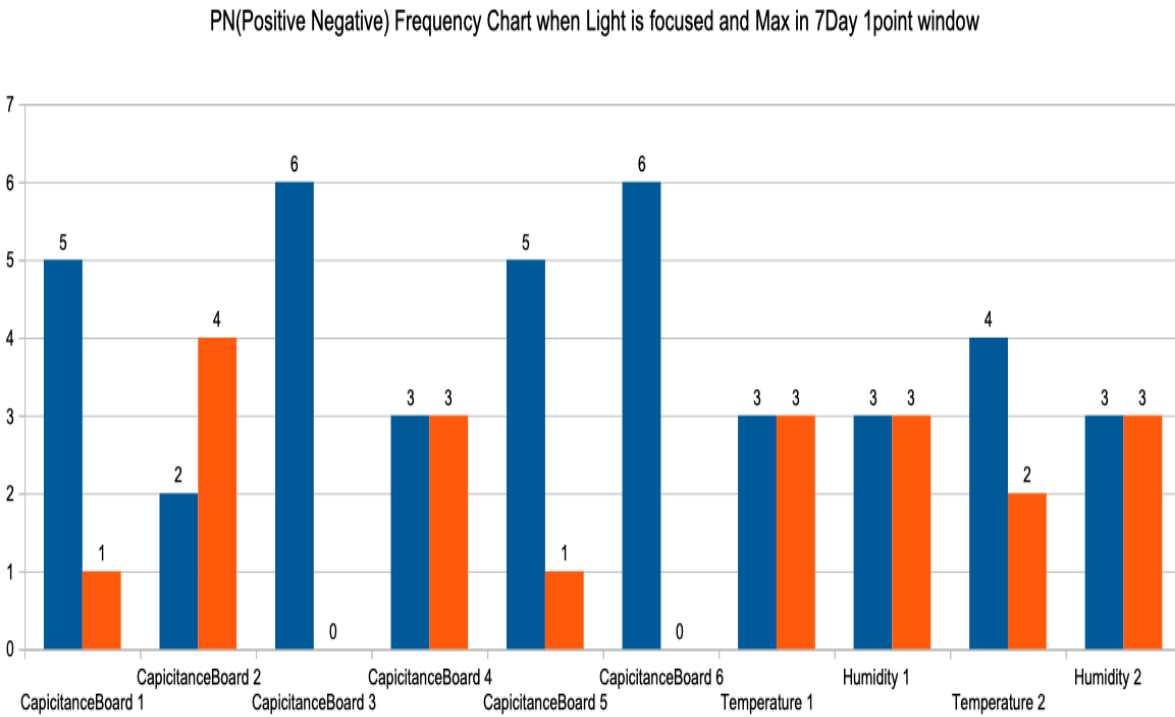
Figure 7.1- PN frequency chart when light is the focus and maximum in 7-day-1-sample window. Positive is blue and negative is red.

Figure 7.2- PN frequency chart when light is the focus and maximum in 7-day-16-samples window. Positive is blue and negative is red.

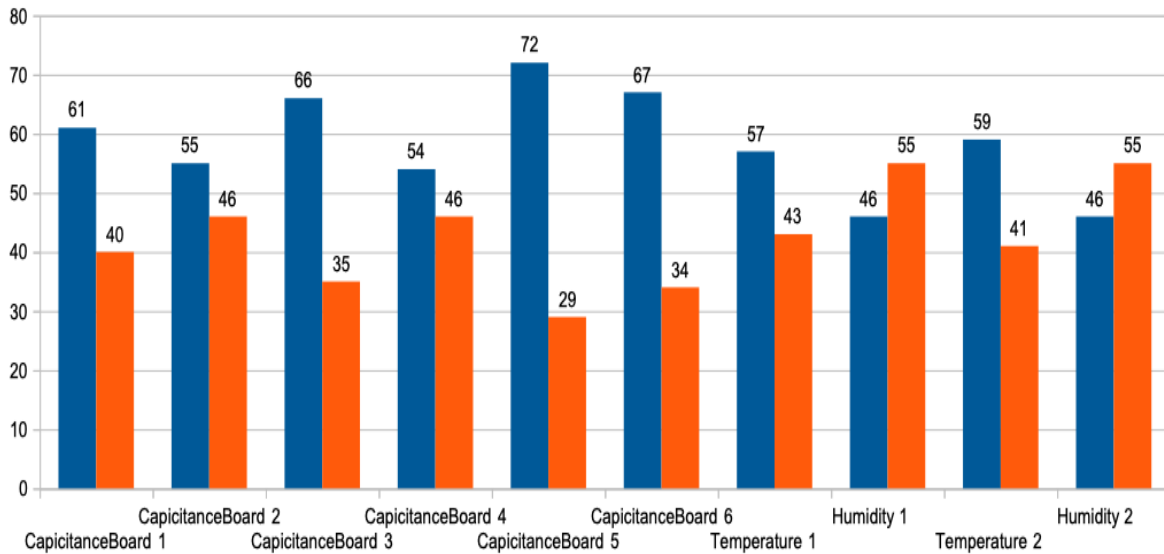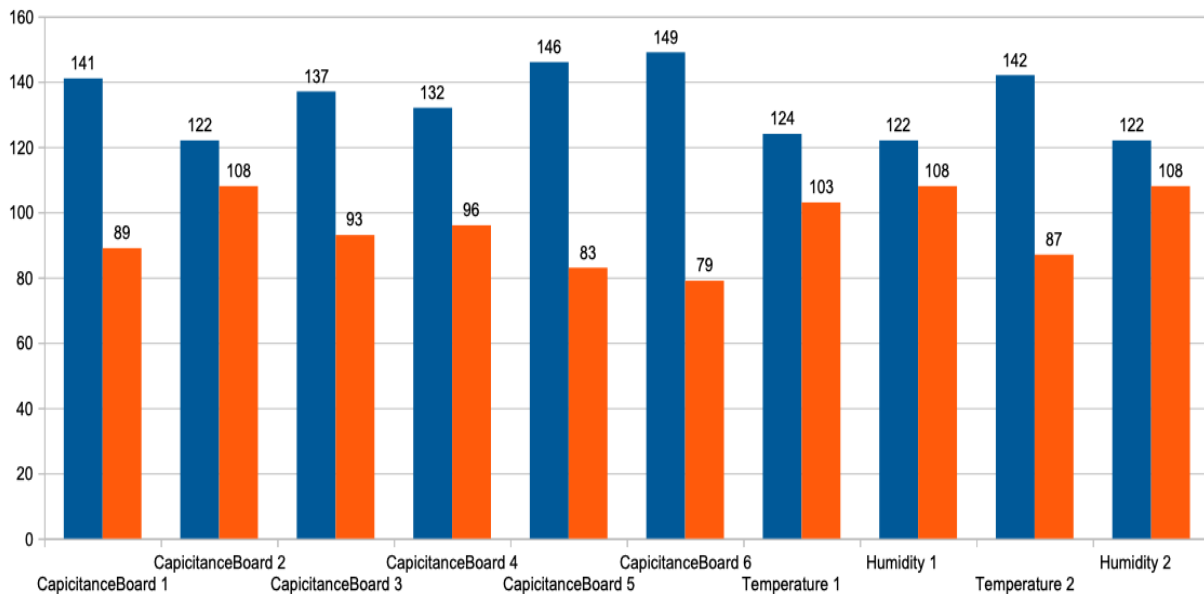*Figure 7.3- PN frequency chart when light is the focus and maximum in 3-day-16-samples window. Positive is blue and negative is red.*

7.2.2. PN charts focusing on maximum light at nighttime

For these nighttime PN charts for light, it should be noted that the light is mostly constant (mostly 0 value) due to the darkness. And so, the information on the charts based on light as the focus is not conclusive. For example, in the day results of the previous section, more light could indicate more sunshine and higher temperature, but at night linkage between light and temperature is not clear.

Figure 7.4 and 7.5 shows that in most of the capacitance boards there is a dominant increase in capacitance. Only in a few cases of capacitance board 1 and capacitance board 4 of Figure 7.4, and capacitance board 3 of Figure 7.5, showed a dominant decrease in activity. Also, considering figures 7.4 and 7.5, capacitance boards 2, 5, and 6 show consistent increase.

An interesting point about both figures is that most activity is on capacitance board 5 (Figure 7.6), which is our capacitance board having the most honey on it. This may show that even at night bees are busy and active on capacitance boards with honey. Or perhaps the bees are clustering on the board. Both figures also show larger two-night activity or clustering on capacitance board 2 (Figure 7.7), which is the capacitance board in the brood nest that has more comb structures built on it.

Figure 7.4- PN (positive negative) frequency chart when light is the focus and maximum in 7-night 6-sample window. Positive is blue and negative is red.
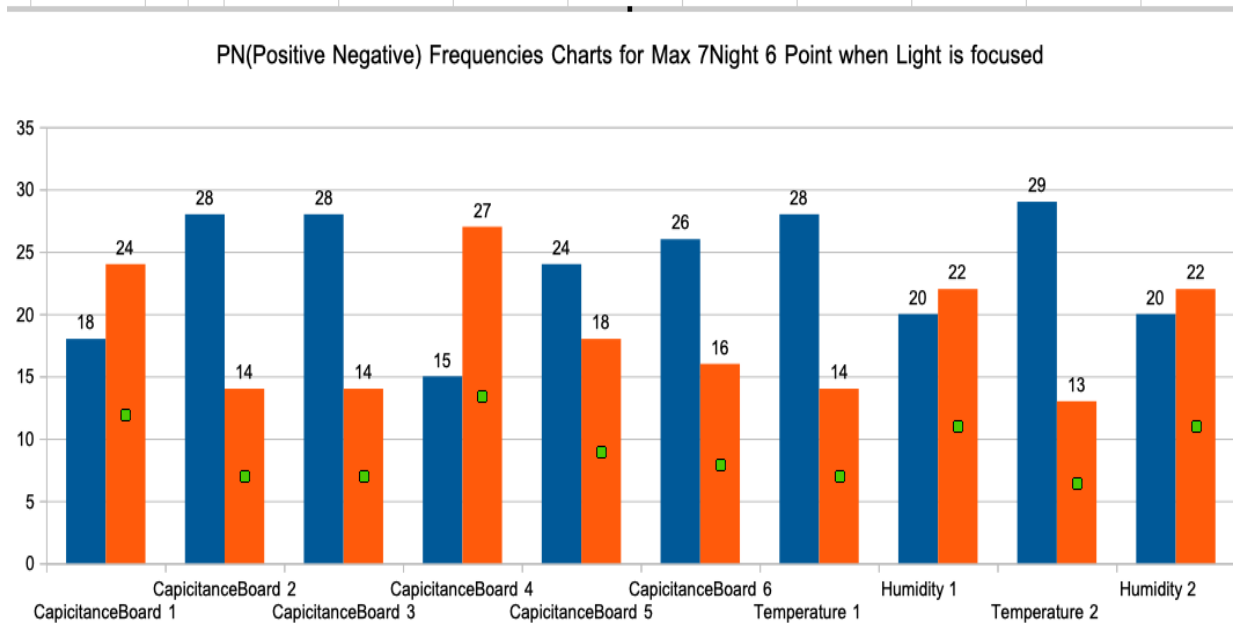

Figure 7.5- PN (positive negative) frequency chart when light is the focus and maximum in 3-night 6-sample window. Positive is blue and negative is red.

Figure 7.6- Capacitance board 5, October 3, 2018. The capacitance board is built by Valerie Beynon [2].

Figure 7.7- Capacitance board 2, October 3, 2018. The capacitance board is built by Valerie Beynon [2].

7.2.3. PN charts focusing on minimum light at daytime

Consider figures 7.9, 7.9 and 7.10. The interesting point about these three figures are that for all of them regardless of the window size, there is a decrease in the activity in most capacitance boards when the light is minimum, except for capacitance board 2 and capacitance board 5 which shown an increase. Especially capacitance board 5, where the increase is more than capacitance board 2. This may also be a sign that when the light is low the bees are more active on capacitance board 5 which has the most honey, and capacitance board 2 the most comb in the brood box. It might also mean they are clustering or are doing other activities.

Figure 7.8- PN (positive negative) frequency chart when light is the focus and minimum in 7-day-1-sample window. Positive is blue and negative is red.

Figure 7.9- PN (positive negative) frequency chart when light is the focus and minimum in 7-day-16-samples window. Positive is blue and negative is red.
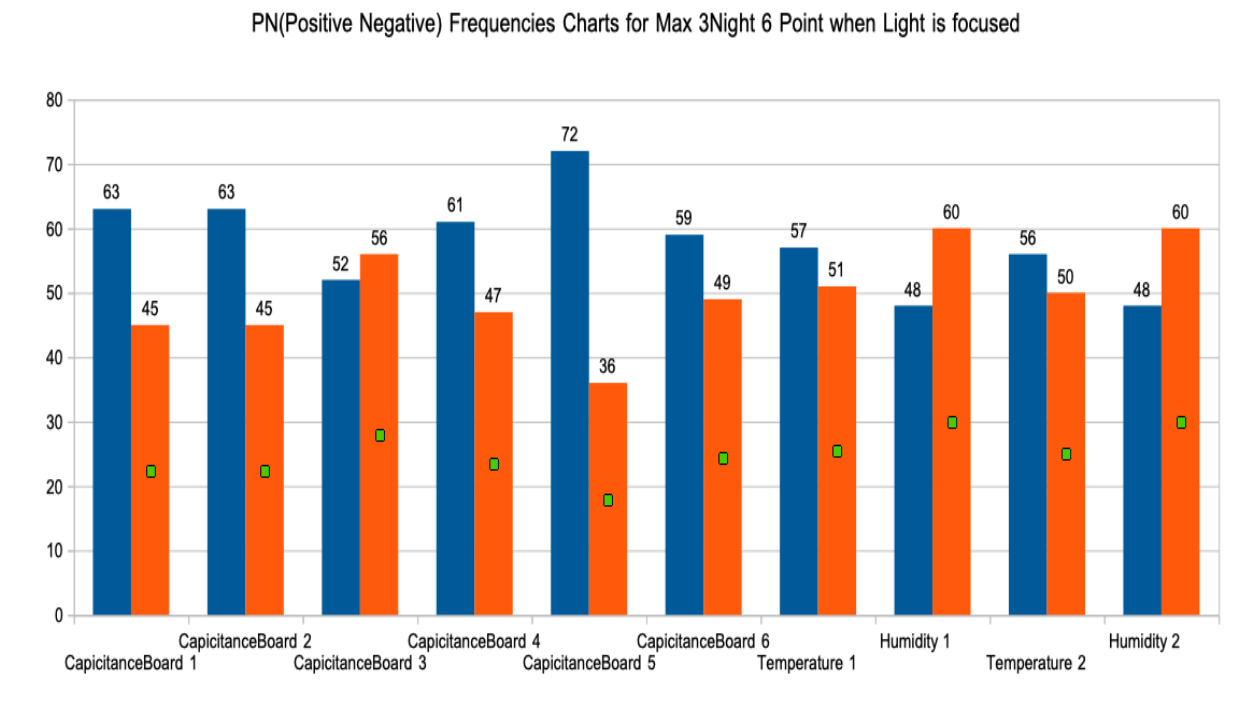

Figure 7.10- PN (positive negative) frequency chart when light is the focus and minimum in 3-day-16-samples window. Positive is blue and negative is red.

7.2.4. PN charts focusing on minimum light at night

Like section 7.2.2, light levels are mostly 0 at night, and conclusions are difficult to find. But it can be seen in figures 7.11 and 7.12 that capacitance board 2, capacitance board 5, and capacitance board 6 that bee activity or clustering were constantly increasing in both window sizes of 3 and 7.
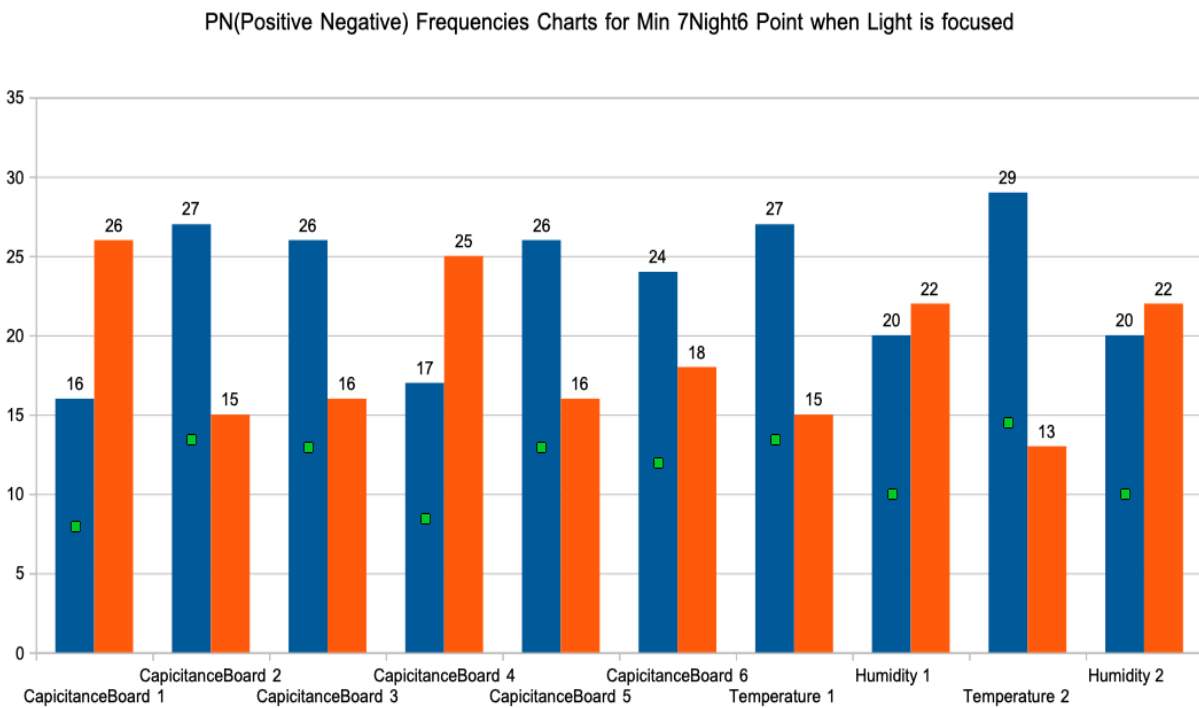
Figure 7.11- PN (positive negative) frequency chart when light is the focus and minimum in 7-night 6-ssample window. Positive is blue and negative is red.

PN(Positive Negative) Frequencies Charts for Min 3Night 6 Point when Light is focused
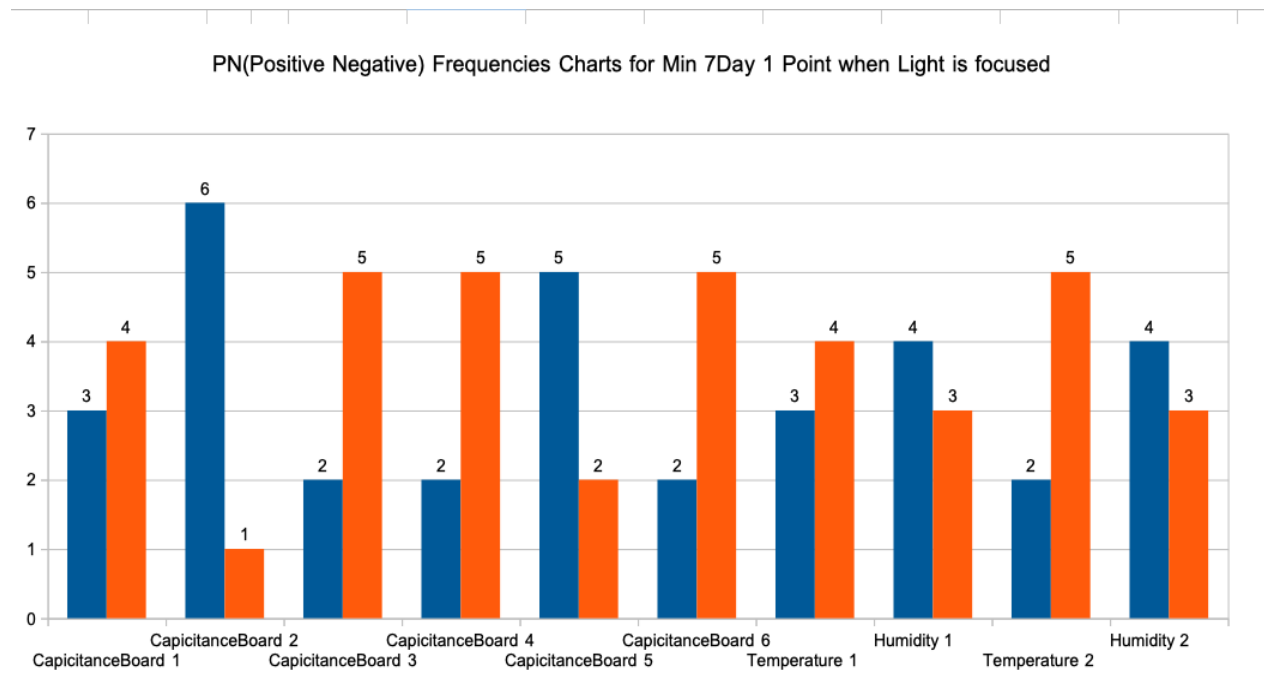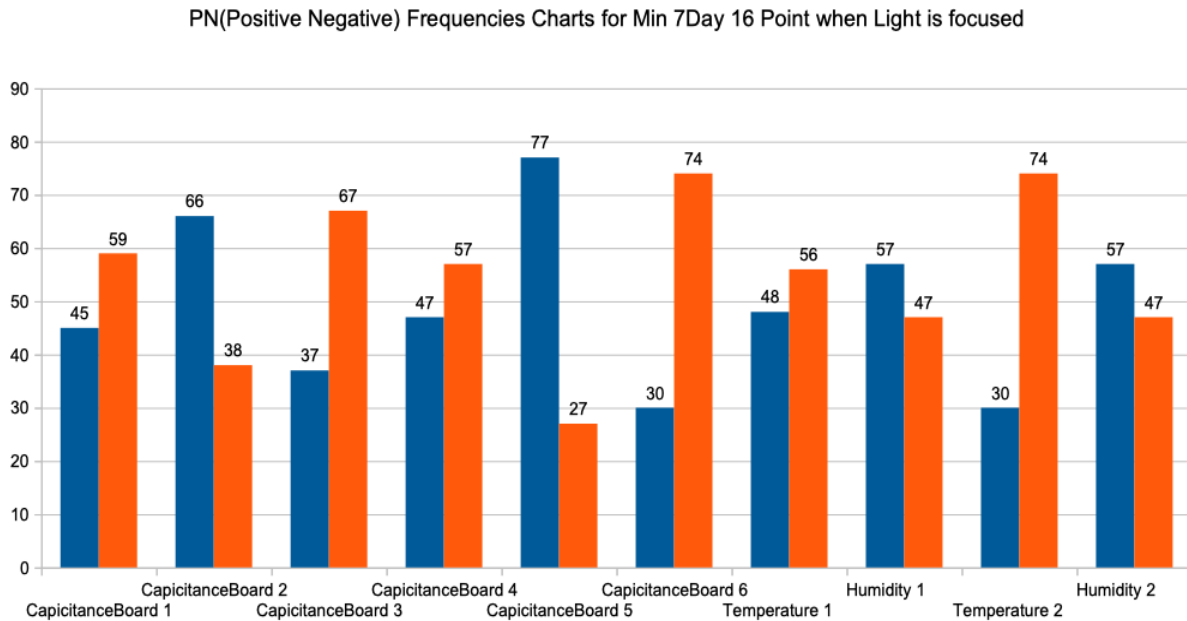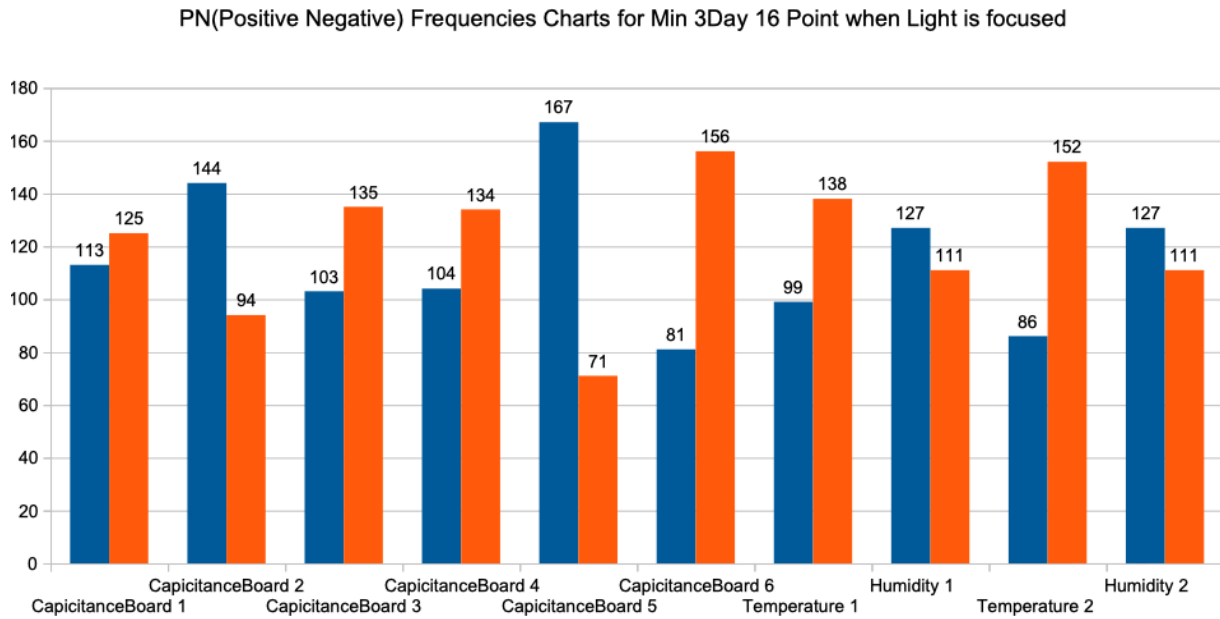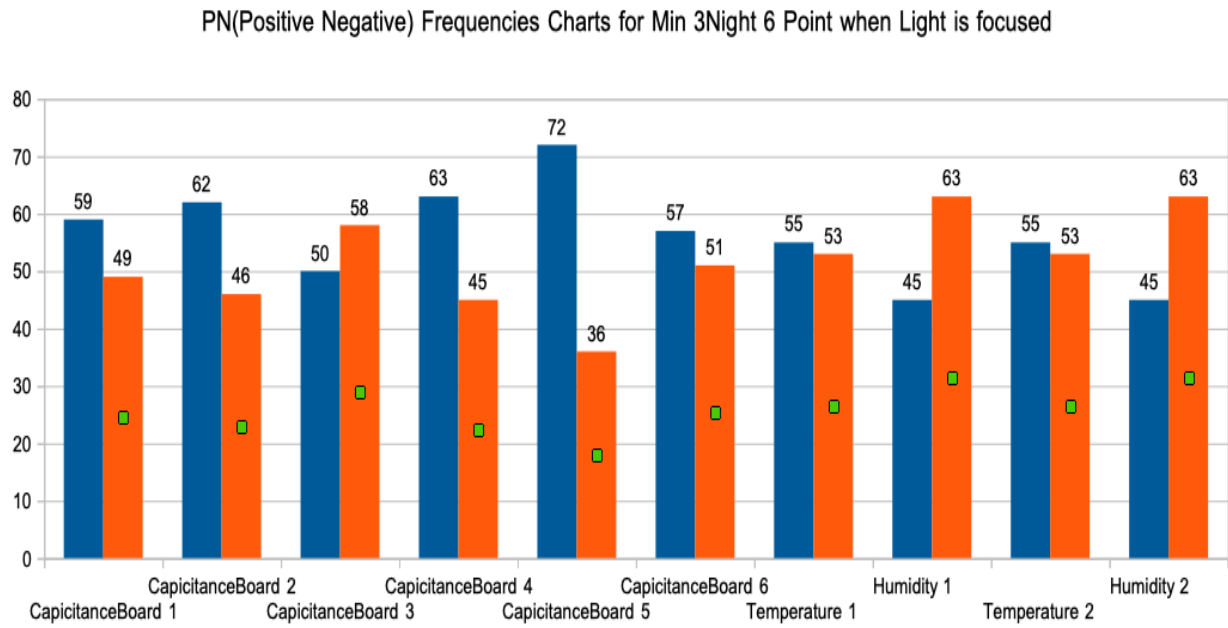
Figure 7.12- PN (positive negative) frequency chart when light is the focus and minimum in 3-night 6-sample window. Positive is blue and negative is red.

7.3. Conclusions of PN charts in Chapter 7

The PN charts shown in this chapter are a good way to see the total number of increase and decrease occurrences on all sensors for when a given factor is a focus. In this chapter, capacitance board activities when the light was the focus, showed how the light effects activities on the capacitance boards, with most affect during the day. During the night as the light was in most cases 0, making comparisons was difficult.

Overall, whether the light is minimum or maximum the bees are usually having more activities on the brood capacitance board 2 and on the honey capacitance board 5, which are the capacitance boards that have significant comb on them. It is also important to note that both these capacitance boards were in the middle of their hive boxes.

Table 7.2 summaries the results of the various analysis in this chapter. It can be seen that, the maximum light during the day for the windows sizes with 16-samples given the same dominant changes in sign in all capacitance boards. In the same way, all minimum light window sizes give the same result. Considering nighttime, as the light would be mostly zero, the result of dominant frequency change for maximum are the same as the dominant change for minimum with some differences in frequencies.

Table 7.2. PN light sensor charts result summary for various window size and points.

Window size & period	Obtained frequencies of changes			
	Global frequency change between capacitance boards	Common dominant frequency change sign between capacitance boards	Specific dominant frequency change sign of capacitance boards (not common)	Figure number
maximum light 7-day-1-sample	No (not enough data points in a day was reviewed)	Yes	Yes	7.1
maximum light 7-day-16-samples	Yes	Yes	-	7.2
maximum light 3-day-16-samples	Yes	Yes	-	7.3
minimum light 7-day-1-sample	-	Yes two common group	-	7.8
minimum light 7-day-16-samples	-	Yes two common group	-	7.9
minimum light 3-day-16-samples	-	Yes two common group	-	7.10
maximum light 7-night 6-sample	-	Yes two common group	-	7.4
maximum light 3-night 6-sample	-	Yes	Yes for one capacitance board	7.5
minimum light 7-night 6-sample	-	Yes two common group	-	7.11
minimum light 3-night 6-sample	-	Yes	Yes for one capacitance board	7.12

7.3.1. Conclusion of comparing minimum and maximum of day in PN charts

When the light is maximum the hive activity in all capacitance boards increased; but when it is minimum the activity in all of them, except in capacitance boards 2 and 5, is decreased.

7.3.2. Conclusion of comparing minimum and maximum of night in PN charts

As mentioned in section 7.2.2, the measured light level was zero, and for both minimum and maximum for 7-Night and 3- Night capacitance boards 2, 5, and 6 have their dominant frequency change increasing. However, for 7- Night windows capacitance boards 1 and 4, and for the 3-Night window capacitance board 3, the dominant frequency change was decreasing.

Chapter 8. Conclusions

In this thesis various software programs were written for analyzing the data obtained from various hive system. The main section of this thesis focused on the data analysis to find frequent relationships between the sensor data. This was investigated over various sliding window sizes in days and nights. These rules were obtained based on searching for relationships between factors, and when the focused factor was maximum or minimum. Two methods were used in this thesis for the data obtained in 2018 including, All-Factors Relationship frequent rules (shown on the All-Factors-Relationships charts) and Two-Factors-Relationship (shown by PN frequencies and charts).

Comparing 3-day and 7-day windows for All-Factors-Relationships rules revealed the following results (shown in detail in Appendix E). It can be seen that the 7-day-1-sample analysis generally did not have enough samples for extracting generalizing rules. By comparing the maximum day analysis using the 3-day-16-sample and the 7-day-16-sample, it can be seen that having the 16 samples per day gave more times in a day when the data was analyzed, and so this gives beneficial results. The 3-day-16-sample windows made the relationship more apparent and more suitable for finding the dominant common rules. However, a few specific rules could still be obtained from 7-day-16-sample windows. Regarding minimum activity situations for both 7-day and 3-day sampling, It was found that they both give good results. For example, both show capacitance board 1 and capacitance board 4 activity having a minimum activity with temperature drop.

The results obtained from the All-Factors-frequent rules showed some common and frequent rules between some capacitance boards in some periods. It also showed some global rules between

all capacitances in some periods. Considering maximum-rules during day, the 3-day-16-sample window method resulted in several common rules for both maximum and minimum rules. Considering global maximum frequent rules at night for both 3-Night-6-Samples, and 7-Night-6-Samples is "SPPPPPPPPP", that is showing increase in all sensors' amount together except for 7-Night-6-Samples rules of light sensor and Capacitance Board3. On the other hand, "SNNNNNNNNNN" is the global most frequent rule for minimum rules, that is showing decrease in all sensors amount for both 3-Night-6-Samples, and 7-Night-6-Samples except light in 7-Night-6-Samples.

The results obtained from PN charts in 7-day-16-samples and 3-day-16-samples were showing that the dominant changes were mostly globally increasing for most capacitance boards activity when light was maximum. The other periods when light was minimum or maximum, most sensors amount did not show globally dominant increasing or decreasing changes. when Light is minimum the activity decreased in all of them, except in capacitance boards 2 and 5. As mentioned in section 7.2.2, at night the measured light level was zero. Also, for both minimum and maximum for 7-Night and 3-Night capacitance boards 2, 5, and 6 have their dominant frequency change increasing. However, for 7-Night windows capacitance boards 1 and 4, and for the 3-Night window capacitance board 3, the dominant frequency change was decreasing.

It is found that the All-Factors-Relationships' rules, which looks for all relationships between all the factors to a focused factor, is more reliable than Two-Factors-Relationships. This is because the Two-Factors-Relationships only consider one other factor's relationship to the focused factor.

Chapter 9. Future Work

There are several things that could be considered for future study based on the work done in this thesis. Some future works could be:

- The data was stored on SD card, and it needed removal from board for reading its data. So, implementing a wireless communication would be a benefit so the card does not need to be removed.

- Redesign the circuits and boards so that disconnection of some boards will not lead to wrong data and time.

- Try to use smaller components for the hardware, to enable a smaller size to fit in water resistant setup.

- Add internet connection and design a mobile application so the data can be accessed remotely.

- Add other components like bee counters, camera, etc.

- Implement predictive machine learning methods for the raw hive data to give future predictions.

- Hardware could be powered off when not actively taking senor readings, to both decrease the power consumption and decrease any possible effect on the bees in the hive boxes.

- Using more temperature sensors especially for each board so the effect of temperature would become more apparent.

References

1. Dr. Currie research group web based system data obtained from hive at University of Manitoba: Department of Entomology - Dr. R.W. Currie (umanitoba.ca), Accessed 22:14 CST May 9, 2021.

2. Valerie Beynon M.Sc. Thesis, University of Manitoba, "Cell-level temperature and hive state monitoring for a remote smart beehive system," (2018).

3. https://en.m.wikipedia.org/wiki/Honey, Accessed 22:19 CST May 9, 2021.

4. https://en.m.wikipedia.org/wiki/Honey_bee, Accessed 22:22 CST May 9, 2021.

5. Simpson, James. "Nest climate regulation in honey bee colonies," Science 133.3461 (1961): 1327-1333.

6. Parsons, Graham. "Factors affecting the overwintering dynamics and structure of honey bee (Apis mellifera) populations," (2017).

7. https://www.arnia.co.uk, Accessed 20:01 CST April 25, 2021.

8. http://solutionbee.com, Accessed 20:31 CST April 29, 2021.

9. https://www.kickstarter.com/projects/970162847/beepi-a-multisensor-electronic-Beehive-monitor, Accessed April 25 2021 23:37 CST.

10. https://www.osbeehives.com, Accessed 20:01 CST April 25, 2021.

11. https://www.3bee.it/en/hive-tech, Accessed 23:38 CST April 25, 2021.

12. https://www.indiegogo.com/projects/apis-tech-smart-hive-monitor-for-every-Beehive, Accessed May 18, 2021.

13. https://broodminimumder.com, Accessed May 18, 2021.

14. https://pollenity.com/product/beebot, Accessed Nov 13, 2020.

15. https://pollenity.com/product/uhive, Accessed Nov 13, 2020.

16. https://www.mitacs.ca/en/impact/sfu-lab-and-chilliwack-company-are-buzzing-tech-advances, Accessed May 18, 2021.

17. https://phys.org/news/2017-08-technology-tracks-bee-honey-health.html, Accessed May 19, 2021.

18. https://ca.sports.yahoo.com/news/smart-beehive-reduce-labor-goes-211058432.html, Accessed May 18, 2021.

19. https://www.educba.com/supervised-learning-vs-unsupervised-learning, Accessed 20:27 CST October 9, 2020.

20. https://simplicable.com/new/supervised-learning, John Spacey, Accessed December 2020.

21. https://theappsolutions.com/blog/development/unsupervised-machine-learning, Accessed December 11, 2020.

22. https://www.talend.com/resources/data-mining-techniques, Accessed 20:10 CST October 9, 2020.

23. https://www.digitalvidya.com/blog/data-cleaning-techniques, Accessed Jan 08, 2021.

24. https://elitedatascience.com/data-cleaning, Accessed Jan 08, 2021.

25. https://www.kdnuggets.com/2016/12/data-science-basics-types-patterns-mined-data.html, Accessed Jan 08, 2021

26. https://builtin.com/data-science/10-data-mining-techniques-data-scientists-need-their-toolbox, by James Le, Accessed 11:44 CST November 9, 2020.

27. http://www.kdd.org/kdd2016/topics/view/frequent-pattern-mining#_blank, Accessed 22:40 CST May 9, 2021.

28. https://riptutorial.com/algorithm/example/25071/sliding-window-algorithm-basic-information, Accessed 19:10 CST May 22, 2019.

29. https://www.farnell.com/datasheets/1682209.pdf, Accessed April 29, 2021.

30. Adafruit Si7021 Temperature & Humidity Sensor Breakout Board [STEMMA QT] : ID 3251 : Adafruit Industries, Unique & fun DIY electronics and kits. https://www.adafruit.com/product/3251#_blank, Accessed 22:12 CST May 9, 2021.

31. https://learn.adafruit.com/adafruit-data-logger-shield, Accessed 22:34 CST May 9, 2021.

32. TCA9548A I2C Multiplexer : ID 2717 Adafruit Industries, Unique & fun DIY electronics and kits. https://www.adafruit.com/product/2717, Accessed 22:09 CST May 9, 2021.

33. https://superuser.com/474530/excel-delete-row-if-cell-in-certaincolumn-is-blank, Accessed October 10, 2017.

34. Overview | Adafruit APDS9960 breakout | Adafruit Learning System - https://learn.adafruit.com/adafruit-apds9960-breakout, Accessed 1:23 Am CST July 13, 2021.

35. Overview | Adafruit MPR121 12-Key Capacitive Touch Sensor Breakout Tutorial | Adafruit Learning System https://learn.adafruit.com/adafruit-mpr121-12-key-capacitive-touch-sensor-breakout-tutorial/overview, Accessed 22:44 CST May 9, 2021.

Appendix A.   First format of All-Factors-Relationship output for

capacitance board 1

In this section, a sample output for capacitance board 1 that was generated by the software in chapter 6 is shown. The output here does not have minimum frequency limit.

```
Wholecapacitance 1 : PPPPPPPPPPP3
Wholecapacitance 1 : NPPNPPNPPPP4
Wholecapacitance 1 : NPPPPPPPPPP7
Wholecapacitance 1 : PPNPPPPNNPP2
Wholecapacitance 1 : NPPPPPPPNPP4
Wholecapacitance 1 : NPPNPPNNPPP1
Wholecapacitance 1 : PPNPPNPPPPP2
Wholecapacitance 1 : PPNPPPPPNPP2
Wholecapacitance 1 : NNPPNPPNNPP1
Wholecapacitance 1 : PPPNNPPPNNN2
Wholecapacitance 1 : NPPNNPNPNNN2
Wholecapacitance 1 : NPNNPPNPPPP1
Wholecapacitance 1 : PPNNPNNNPPP3
Wholecapacitance 1 : NPNPPPPPPPP2
Wholecapacitance 1 : NPPNNNPPNNN1
Wholecapacitance 1 : PPNNNNNNNPP5
Wholecapacitance 1 : PPPNPNNNNPP1
Wholecapacitance 1 : PPPPPNPNPPP1
Wholecapacitance 1 : PPNNPPNNPPP4
Wholecapacitance 1 : NPPNNNPPPNN1
Wholecapacitance 1 : PPNPPPNPPPP1
Wholecapacitance 1 : PPPPPNPPNPP1
Wholecapacitance 1 : PPNPPPPPPPP3
Wholecapacitance 1 : PPPPPPPPNPP2
Wholecapacitance 1 : PPNPPNNPNNN1
Wholecapacitance 1 : PNPNNPPNNNN1
Wholecapacitance 1 : PPPNPPPPPPP7
Wholecapacitance 1 : NPPPPPPPPNN4
Wholecapacitance 1 : PPPPNNPNNNN1
Wholecapacitance 1 : NPNPPPPNPPP3
Wholecapacitance 1 : PPNNPPNNNPP3
Wholecapacitance 1 : PNPPNPNPNNN1
Wholecapacitance 1 : NPPNPNNNPPP1
Wholecapacitance 1 : NPNPNNPPPPP1
Wholecapacitance 1 : NPNPNPPPNPP1
Wholecapacitance 1 : NNPPNPPPNNN3
Wholecapacitance 1 : PPPNPPPPNNN1
Wholecapacitance 1 : PPPNNPNPPNN2
Wholecapacitance 1 : PPPPNPNNNPP1
```

```
Wholecapacitance 1 : PPNPPNNNPNN2
Wholecapacitance 1 : NPPNPPPPPPP6
Wholecapacitance 1 : NNPPNNPPNNN1
Wholecapacitance 1 : NPNNPPPNNNN1
Wholecapacitance 1 : NPNNPPNNNPP2
Wholecapacitance 1 : PNPPNPPPNNN1
Wholecapacitance 1 : PPPNPNNNPPP1
Wholecapacitance 1 : NPPPNNSPNPP1
Wholecapacitance 1 : PPNNNPPPNPP1
Wholecapacitance 1 : PPPNPPNPPNN1
Wholecapacitance 1 : PPNNPPPPPPP1
Wholecapacitance 1 : PPPPNPPNNPP5
Wholecapacitance 1 : PPPNPPNPPPP2
Wholecapacitance 1 : NPPNPNPNNNN1
Wholecapacitance 1 : NPPNNPNPPPP2
Wholecapacitance 1 : PPNPPPPPPNN3
Wholecapacitance 1 : NPPPPPNPNPP1
Wholecapacitance 1 : NPNNNPNNNPP4
Wholecapacitance 1 : NNPPNNPPPNN2
Wholecapacitance 1 : PPPNNPNNNPP3
Wholecapacitance 1 : PPPPNPNPNNN2
Wholecapacitance 1 : NNPPNPPNNNN1
Wholecapacitance 1 : PPNNNPPPNNN1
Wholecapacitance 1 : PPNPPNPNPNN2
Wholecapacitance 1 : PPPNNPPPPPP1
Wholecapacitance 1 : PPPPNNNPPNN2
Wholecapacitance 1 : NPPNPPPPPNN1
Wholecapacitance 1 : NPPNPNPNPNN1
Wholecapacitance 1 : NPPPPNNPNPP1
Wholecapacitance 1 : NPPNNPNNNPP5
Wholecapacitance 1 : NPNPPNPPPPP1
Wholecapacitance 1 : NNPNNPPNNNN1
Wholecapacitance 1 : PPPNNPPNPNN1
Wholecapacitance 1 : PPPNNPPPPNN1
Wholecapacitance 1 : PPPPPNNNPNN1
Wholecapacitance 1 : NPPPNPPPNPP1
Wholecapacitance 1 : PPPNPPPNPPP1
Wholecapacitance 1 : NPPNNNNPNPP1
Wholecapacitance 1 : NPNNPNNNNPP3
Wholecapacitance 1 : NPNNPPNNPPP2
Wholecapacitance 1 : NNPPNNNPNPP1
Wholecapacitance 1 : PPPNPPNNNPP1
Wholecapacitance 1 : PPNPNNPPPNN2
Wholecapacitance 1 : NPPNPPNPNNN1
Wholecapacitance 1 : NNNPPPNNNNN1
Wholecapacitance 1 : NPPNPPPNPPP1
Wholecapacitance 1 : PNPPPPNPNNN1
Wholecapacitance 1 : SPPPPPNNPPP1
Wholecapacitance 1 : PPPNNPPPNPP1
Wholecapacitance 1 : NPPPNNNNNNN1
Wholecapacitance 1 : SSSSSSSSSSS5
Wholecapacitance 1 : NPPNNNNPNNN1
Wholecapacitance 1 : PPNNPNNPNPP1
Wholecapacitance 1 : PPNNNPNNPPP3
```

```
Wholecapacitance 1 : PPPNPNNPPNN1
Wholecapacitance 1 : PPPPNNPPPPP1
Wholecapacitance 1 : PPPNPPPPNPP2
Wholecapacitance 1 : NNNPNPPNNNN1
Wholecapacitance 1 : PPPPPPNNPNN1
Wholecapacitance 1 : PPPPNPNPPNN4
Wholecapacitance 1 : NPPPPPPNNNN1
Wholecapacitance 1 : NPNPNNNNNNN4
Wholecapacitance 1 : PPNPNNNPNNN1
Wholecapacitance 1 : NPPNPPPPNPP4
Wholecapacitance 1 : NNNNNPNNNPP1
Wholecapacitance 1 : NPNNNNNNNPP2
Wholecapacitance 1 : NPNPPNPPNPP2
Wholecapacitance 1 : NPNPPPPPPNN2
Wholecapacitance 1 : NPPPPNPNNPP1
Wholecapacitance 1 : PPNPPPPNPPP2
Wholecapacitance 1 : NPPNPNPPPNN1
Wholecapacitance 1 : PPPNPPNPNPP2
Wholecapacitance 1 : NNPNNPNNNPP3
Wholecapacitance 1 : PPPPPPNPPPP1
Wholecapacitance 1 : PNNNPPPPNNN1
Wholecapacitance 1 : NPPNNNNNPPP1
Wholecapacitance 1 : PNNPPPPPPNN1
Wholecapacitance 1 : PPNNPPNNPNN1
Wholecapacitance 1 : PPNNNPNNSPP1
Wholecapacitance 1 : PPPNNNNPPNN1
Wholecapacitance 1 : NPPPNNNPNNN1
Wholecapacitance 1 : PPPPNNPPPNN4
Wholecapacitance 1 : NNPNPPNNPPP1
Wholecapacitance 1 : NNNNNPPNNPP2
Wholecapacitance 1 : PPPPPPNPPNN3
Wholecapacitance 1 : NPNNPNPNNPP1
Wholecapacitance 1 : NPPNNNNNNPP2
Wholecapacitance 1 : NPPNPNPPNPP1
Wholecapacitance 1 : PNNNNPNNPPP1
Wholecapacitance 1 : PPPNNPSPPNN1
Wholecapacitance 1 : PPNPPNPNPPP2
Wholecapacitance 1 : NPPPNPNPPNN1
Wholecapacitance 1 : PNPNPPPPPNN1
Wholecapacitance 1 : NNNNPPPNNPP1
Wholecapacitance 1 : NPNNNNPNPPP1
Wholecapacitance 1 : PPNNNNNNNNN4
Wholecapacitance 1 : PPPNNNNPNPP1
Wholecapacitance 1 : PPNNPPPNNPP3
Wholecapacitance 1 : PNNNNPNNNPP2
Wholecapacitance 1 : PNPPNPPPPNN1
Wholecapacitance 1 : NPPPNNNPPNN2
Wholecapacitance 1 : NPNNPPPPNPP1
Wholecapacitance 1 : PPPPPPNSNPP1
Wholecapacitance 1 : NPPPNPNNNPP1
Wholecapacitance 1 : NPNPPNPNNNN1
Wholecapacitance 1 : PNPNNPNNPNN1
Wholecapacitance 1 : PNPPPPPPPNN1
Wholecapacitance 1 : PPPPPPPNNPP1
```

116

```
Wholecapacitance 1 : NPPNPPNPNPP1
Wholecapacitance 1 : NPPPNPPPPNN1
Wholecapacitance 1 : NNPNNPPNNPP1
Wholecapacitance 1 : NPNPNPNNNPP2
Wholecapacitance 1 : NNPNNNNNPNN1
Wholecapacitance 1 : NPPNNNNNNNN1
Wholecapacitance 1 : PNNNSPNNNPP1
Wholecapacitance 1 : NPNPPNPNPPP1
Wholecapacitance 1 : NNPPPPPNPPP1
Wholecapacitance 1 : NPNPNPPNNPP1
Wholecapacitance 1 : PPNPPPPPNNN1
Wholecapacitance 1 : NNNNNNNNPNN1
Wholecapacitance 1 : NPPNPPNNNPP1
Wholecapacitance 1 : PPPNNNNNNPP2
Wholecapacitance 1 : NPNNPPPNNPP1
Wholecapacitance 1 : PNNNNPPNNPP1
Wholecapacitance 1 : NPNPPPPNNPP1
Wholecapacitance 1 : NPNPPNPNPNN1
Wholecapacitance 1 : NPPNPPPNNPP1
Wholecapacitance 1 : NPPPPPNPPNN1
Wholecapacitance 1 : PNPNNNNNPNN1
```

Appendix B.   Second format of All-Factors-Relationship output for capacitance board 1

In this section, a sample output for capacitance board 1 that was generated by the software in Chapter 6 is shown. The output has a comma added between rules and their frequency for extraction to Excel.

```
Wholecapacitance 1 , PPPPPPPPPPP,2
Wholecapacitance 1 , NPPNPPNPPPP,4
Wholecapacitance 1 , NPPPPPPPPPP,6
Wholecapacitance 1 , PPNPPPPNNPP,2
Wholecapacitance 1 , NPPPPPPPNPP,4
Wholecapacitance 1 , NPPNPPNNPPP,1
Wholecapacitance 1 , PPNPPNPPPPP,3
Wholecapacitance 1 , PPPPPPPPNPP,3
Wholecapacitance 1 , PPNPPPPPPPP,2
Wholecapacitance 1 , NPNNNPNPPNN,1
Wholecapacitance 1 , NNPNNPNPNNN,1
Wholecapacitance 1 , NPNNPPNPPPP,1
Wholecapacitance 1 , NPPPNPPPPNN,2
Wholecapacitance 1 , NPNNPNNNNNN,1
Wholecapacitance 1 , PPPNNNPPNNN,1
Wholecapacitance 1 , PPNNNNNNNPP,7
Wholecapacitance 1 , PPPNPNNNNPP,1
Wholecapacitance 1 , PPPPPNPNPPP,1
Wholecapacitance 1 , PPNNPPNNPPP,4
Wholecapacitance 1 , NPPNNNNPPNN,1
Wholecapacitance 1 , PPNPNNPPPPP,1
Wholecapacitance 1 , PPPPPNPPPNPP,1
Wholecapacitance 1 , NPPNNPNNPNN,1
Wholecapacitance 1 , NPPNNPNPNNN,2
Wholecapacitance 1 , NPNPNPNNNPP,3
Wholecapacitance 1 , PPNNPNNNPPP,2
Wholecapacitance 1 , PPPNPPPPPPP,5
Wholecapacitance 1 , PNPNNNNPNNN,1
Wholecapacitance 1 , NPNPPPPNPPP,2
Wholecapacitance 1 , PPNNPPNNNPP,3
Wholecapacitance 1 , PNPPNPNPNNN,1
Wholecapacitance 1 , NPPNPNNNPPP,1
Wholecapacitance 1 , NPNPNNPPPPP,1
Wholecapacitance 1 , NPNPNPPPNPP,1
Wholecapacitance 1 , NNPPNPPPNNN,3
Wholecapacitance 1 , PPPNPPPPNNN,1
Wholecapacitance 1 , PPPNNPNPPNN,2
```

```
Wholecapacitance 1 , PPPPNPNNNPP,1
Wholecapacitance 1 , PPNPPNNNPNN,2
Wholecapacitance 1 , NPPNPPPPPPP,6
Wholecapacitance 1 , NNPPNNPPNNN,1
Wholecapacitance 1 , PPPPNNNNNNN,3
Wholecapacitance 1 , NPNNPPNNNPP,2
Wholecapacitance 1 , PNPPNPPPNNN,1
Wholecapacitance 1 , PPPNPNNNPPP,1
Wholecapacitance 1 , NPPPNNSPNPP,1
Wholecapacitance 1 , PPNNNPPPNPP,1
Wholecapacitance 1 , PPPNNPPPNNN,1
Wholecapacitance 1 , PPPNPPNPPNN,1
Wholecapacitance 1 , PPNNPPPPPPP,1
Wholecapacitance 1 , PPPPNPPNNPP,5
Wholecapacitance 1 , PPPNPPNPPPP,2
Wholecapacitance 1 , NPPPNNPNNNN,1
Wholecapacitance 1 , NPPNNPNPPPP,1
Wholecapacitance 1 , PPNPPPPPPNN,2
Wholecapacitance 1 , NPPPPPNPNPP,1
Wholecapacitance 1 , NPNNNPNNNPP,5
Wholecapacitance 1 , NNPPNNPPPNN,2
Wholecapacitance 1 , PPPNNPNNNPP,3
Wholecapacitance 1 , PPPPNPNPNNN,2
Wholecapacitance 1 , NNPPNPPNNNN,1
Wholecapacitance 1 , PPNNNPPPNNN,1
Wholecapacitance 1 , PPNPPNPNPNN,2
Wholecapacitance 1 , PPPNNPPPPPP,1
Wholecapacitance 1 , NNPPNNNPNNN,2
Wholecapacitance 1 , NPNNPPPNPNN,2
Wholecapacitance 1 , NPPPPNNPNPP,1
Wholecapacitance 1 , NPPNNPNNNPP,5
Wholecapacitance 1 , NPNPPNPPPPP,1
Wholecapacitance 1 , NNPNNPPNNNN,1
Wholecapacitance 1 , PPPNNPPNPNN,1
Wholecapacitance 1 , PPPNNPPPPNN,1
Wholecapacitance 1 , PPPPPNNNPNN,1
Wholecapacitance 1 , NPPPNPPPNPP,1
Wholecapacitance 1 , NPPNNNNPNPP,1
Wholecapacitance 1 , NPNNPNNNNPP,3
Wholecapacitance 1 , NPNNPPNNPPP,2
Wholecapacitance 1 , NNPPNNNPNPP,1
Wholecapacitance 1 , PPPNPPNNNPP,1
Wholecapacitance 1 , PPNPNNPPPNN,1
Wholecapacitance 1 , NPPNPPNPNNN,1
Wholecapacitance 1 , NNNPPPNNNNN,1
Wholecapacitance 1 , NPPNPPPNPPP,1
Wholecapacitance 1 , PNPPPPNPNNN,1
Wholecapacitance 1 , SPPPPPNNPPP,1
Wholecapacitance 1 , PPPNNPPPNPP,1
Wholecapacitance 1 , PPPPPPPPPNN,2
Wholecapacitance 1 , NPNNNPPNNPP,1
Wholecapacitance 1 , NPNPNNNNNNN,5
Wholecapacitance 1 , NPPNNNNPNNN,1
Wholecapacitance 1 , PPNNPNNPNPP,1
```

```
Wholecapacitance 1 , PPNNNPNNPPP,3
Wholecapacitance 1 , PPPNPNNPPNN,1
Wholecapacitance 1 , PPPPNNPPPPP,1
Wholecapacitance 1 , PPPNPPPPNPP,2
Wholecapacitance 1 , NNNPNPPNNNN,1
Wholecapacitance 1 , PPPPPPNNPNN,1
Wholecapacitance 1 , PPPPNPNPPNN,4
Wholecapacitance 1 , NPPPPPPNNNN,1
Wholecapacitance 1 , PPPNNNNNNPP,3
Wholecapacitance 1 , PPNPPPPNPPP,2
Wholecapacitance 1 , PPPPNNNPNNN,1
Wholecapacitance 1 , NPPNPPPPNPP,4
Wholecapacitance 1 , NNPNNPNNNPP,3
Wholecapacitance 1 , PNPNNPPNNNN,1
Wholecapacitance 1 , NPPPNNNNNNN,1
Wholecapacitance 1 , NPPNNPSPPPP,1
Wholecapacitance 1 , PPPNPPSPNNN,1
Wholecapacitance 1 , NPNPPPPPPNN,3
Wholecapacitance 1 , PPNNNNNSSNN,1
Wholecapacitance 1 , PNPPPPPPNPP,1
Wholecapacitance 1 , NPPNPNPPPNN,1
Wholecapacitance 1 , NPPPNNNPNNN,2
Wholecapacitance 1 , PPNPNPPPPNN,1
Wholecapacitance 1 , PPPNPPNPNPP,2
Wholecapacitance 1 , PPPPPPNPPPP,1
Wholecapacitance 1 , NNNNNPPNNPP,3
Wholecapacitance 1 , NPPPPPNPPNN,2
Wholecapacitance 1 , NPPNNNNNPPP,1
Wholecapacitance 1 , NNPPNNPPNPP,1
Wholecapacitance 1 , PNNPPPPPPNN,1
Wholecapacitance 1 , PPNNPPNNPNN,1
Wholecapacitance 1 , PPNNNPNNSPP,1
Wholecapacitance 1 , PPPNNNNPPNN,1
Wholecapacitance 1 , PPPPNNPPPNN,4
Wholecapacitance 1 , NPPPPPPPPNN,2
Wholecapacitance 1 , NNPNPPNNPPP,1
Wholecapacitance 1 , PPPPPPNPPNN,3
Wholecapacitance 1 , NPNNPNPNNPP,1
Wholecapacitance 1 , NPNPPNPPNPP,1
Wholecapacitance 1 , NPPNNNNNNPP,2
Wholecapacitance 1 , NPPNPNPPNPP,1
Wholecapacitance 1 , PNNNNPNNPPP,1
Wholecapacitance 1 , PPPNNPSPPNN,1
Wholecapacitance 1 , PPNPPNPNPPP,2
Wholecapacitance 1 , NPPPNPNPPNN,1
Wholecapacitance 1 , PNPNPPPPPNN,1
Wholecapacitance 1 , NNNNPPPNNPP,1
Wholecapacitance 1 , NPNNNNPNPPP,1
Wholecapacitance 1 , PPNNNNNNNNN,4
Wholecapacitance 1 , NPNNNNNNNPP,1
Wholecapacitance 1 , PPPNNNNPNPP,1
Wholecapacitance 1 , PPNNPPPNNPP,3
Wholecapacitance 1 , PNNNNPNNNPP,2
Wholecapacitance 1 , PNPPNPPPPNN,1
```

```
Wholecapacitance 1 , NPPPNNPPPNN,2
Wholecapacitance 1 , NPNNPPPPNPP,1
Wholecapacitance 1 , PPPPPPNSNPP,1
Wholecapacitance 1 , NPPPNPNNNPP,1
Wholecapacitance 1 , NPNPPPNPNNN,1
Wholecapacitance 1 , PNPNNPNNPNN,1
Wholecapacitance 1 , PNPPPPPPPNN,1
Wholecapacitance 1 , PPPPPPPNNPP,1
Wholecapacitance 1 , NPPNPPNPNPP,1
Wholecapacitance 1 , NNPNNPPNNPP,1
Wholecapacitance 1 , PPNPPPPPNPP,1
Wholecapacitance 1 , NNPNNNNNPNN,1
Wholecapacitance 1 , NPPNNNNNNNN,1
Wholecapacitance 1 , PNNNSPNNNPP,1
Wholecapacitance 1 , NPNPPNPNPPP,1
Wholecapacitance 1 , NNPPPPPNPPP,1
Wholecapacitance 1 , NPNPNPPNNPP,1
Wholecapacitance 1 , PPNPPPPPNNN,1
Wholecapacitance 1 , NNNNNNNNPNN,1
Wholecapacitance 1 , NPPNPPNNNPP,1
Wholecapacitance 1 , NPNPPPPPPPP,1
Wholecapacitance 1 , NPNNPPPNNPP,1
Wholecapacitance 1 , PNNNNPPNNPP,1
Wholecapacitance 1 , NPNPPPPNNPP,1
Wholecapacitance 1 , NPNPPNPNPNN,1
Wholecapacitance 1 , NPPNPPPNNPP,1
Wholecapacitance 1 , PNPNNNNNPNN,1
```

Appendix C. **First output for comparison of two parameters from the software in Chapter 6**

Note - Some correction in spelling from the original software was done.

Please Enter the function that you want to perform Enter 'A' for Average, 'S' for start time and end time differences or 'Q' for quit

a

Please choose the first variable for your requested function
Enter 'T' for temperature and Humidity, 'L' for light 'W' for Weight, 'C' for Capacitance-Board

c

Please choose the board number of Capacitance-Board that you want from 1 to 6

2

Please choose the board Part of Capacitance-Board that you want.
if you want the Whole Capacitance-Board(all 12 sensors on it) Average Press W
if you want only one sensor from the board choose its number from 1 to 12.
If you want the outer part near board wire Average (sensors numbers 12 and 1) press L
if you want If you want the outer part far from board wire Average (sensors numbers 6 and 7) press R
if you want the center left two columns Average (sensors numbers 11,10,2,3) press M
if you want the center Right two columns Average (sensors numbers 9,8,4,5) press S

l
copyFvarcapnumorpart: 0

Please choose the Second variable for your requested function Enter 'T' for temperature, 'L' for light 'W' for Weight, 'C' for Capacitance-Board

c

Please choose the board number of Capacitance-Board that you want from 1 to 6

2

Please choose the board Part of Capacitance-Board that you want.
if you want the Whole Capacitance-Board (all 12 sensors on it) Average Press W
if you want only one sensor from the board choose its number from 1 to 12.
If you want the outer part near board wire Average (sensors numbers 12 and 1) press
if you want If you want the outer part far from board wire Average (sensors numbers 6 and 7)
press R
if you want the center left two columns Average (sensors numbers 11,10,2,3) press M
if you want the center Right two columns Average (sensors numbers 9,8,4,5) press S

r

atoi(Svarcapnumorpart)= 0

FvAvgSlopeReadnum:582
SvAvgSlopeReadnum:582

Your First Variable Average is : 18

Your Second Variable Average is : 25

For New Function for your previous period enter 'A'
For changing the time period enter for another Function 'C'
For choosing s new Period Date and new filename for doing another Function press 'F'
For Quit press 'Q'"

Appendix D. **Second output for comparison of two parameters from the software in Chapter 6**

Note - Some correction in spelling from the original software was done.

"Please Enter the function that you want to perform Enter 'A' for Average, 'S' for start time and end time differences or 'Q' for quit

a

Please choose the first variable for your requested function
Enter 'T' for temperature and Humidity, 'L' for light 'W' for Weight, 'C' for Capacitance-Board

t

Please choose the board number of temperature board that you want from 1 to 2

1

Please choose the Second variable for your requested function Enter 'T' for temperature, 'L' for light 'W' for Weight, 'C' for Capacitance-Board

t

Please choose the board number of temperature board that you want from 1 to 2

2

FvAvgSlopeReadnum:291

Temperature Average of the First Variable is : 34

Humidity Average of the First Variable is : 50

Temperature Average of the Second Variable is : 30

Humidity Average of the Second Variable is : 45

For New Function for your previous period enter 'A'
For changing the time period enter for another Function 'C'
For choosing s new Period Date and new filename for doing another Function press 'F'
For Quit press 'Q'

f

***************Getting information of start and end date and time (24 hours format) ************

Please enter start year(number only like 2017 or 17)
20
Please enter start month(number only)
6
Please enter start day(number only)
1
Please enter start hour(number only)
20
Start date and time: 06/01/2020 20:00
Please enter end year (number only like 2017 or 17)
20
Please enter end month (number only)
6
Please enter end day. It should be number only
29
Please enter end hour (number only)
9
End date and time: 06/29/2020 09:00

***Please Enter the text file Name to read the data from (The file Name without extension) ***

new input
******** Hi ,Here is hive activity data *******

***********Here is activity data file: newinput.TXT data********

newinput.TXT
:

****************hive************

****************data************

****************now************

****************finished****************

End of hive data import and formatting them from file. Thanks,
Nadia

Please Enter the function that you want to perform Enter 'A' for Average, 'S' for start time and end time differences or 'Q' for quit

a

Please choose the first variable for your requested function
Enter 'T' for temperature and humidity, 'L' for light 'W' for Weight, 'C' for Capacitance-Board

t

Please choose the board number of Temperature board that you want from 1 to 2

1

Please choose the Second variable for your requested function Enter 'T' for temperature ,'L' for light 'W' for Weight, 'C' for Capacitance-Board

t

Please choose the board number of temperature board that you want from 1 to 2

2
FvAvgSlopeReadnum:1305

temperature Average of the First Variable is : 23

Humidity Average of the First Variable is : 24

Temperature Average of the Second Variable is : 23

Humidity Average of the Second Variable is : 24

For New Function for your previous period enter 'A'
For changing the time period enter For another Function 'C'
For choosing s new Period Date and new filename for doing another Function press 'F'
For Quit press 'Q'

Appendix E.   **Bar charts and explanations of All-Factors-Relationship rules**

Figure E.1 shows the All-Factors-Relationship rules Frequencies' bar charts for maximum 3-day sliding window for 16-samples per day when minimum Frequency >=7. The paragraph number of each description below is marked by the red circles on the related part of Figure E.1. Some of the main visible rules are summarized below.

The rule sentences are defined by "L123456THth" on top line, with the changes of the individual sensors shown below, with Light = L, capacitance sensors =1 to 6, temperature sensor1=T, Humidity1=H, temperature sensor2=t and Humidity2=h. For example:

```
L123456THth
PPPPPPPPPPP
```

1) Capacitance board 2, capacitance board 4 and capacitance board 3 have

```
L123456THth
PPPPPPPPPPP
```

rule as maximum rule as their maximum activity. Also, this rule is one of the maximum light rules. In the same way, capacitance board 6 second most frequent rule is PPPPPPPPPPP as well.

2) On the other hand, the maximum rule of light goes with the maximum

```
L123456THth
PPNPPPPNPPP
```

activity rule of capacitance board 5 and one of maximum of capacitance board 4, which can say that with the reduction of Temperature1 and capacitance board 2 activity, an increase in activity of capacitance boards 4 and 5 happens.

3) The other rule of capacitance board 5 (shown below) shows that when the light has reduced, activity in capacitance board 5 has increased.

```
L123456THth
NPPPPPPPPPP
```

4) The point about the maximum rule of capacitance board 1 which is:

```
L123456THth
PPNNNNNNNPP
```

5) It appears that with an increase in light, Temprature2 and humidity2, there is maximum activity in capacitance board 1 and reduction in other capacitance boards. This may show an movement of bees to capacitance board 1 because of an increase in other boards' temperature. This could be because capacitance board 1 is in the lowest position in the hive and shaded by trees, which can shadow it from hot temperatures of Winnipeg summer sunshine. Maybe this location can be referred as their shelter in hot days.

6) For capacitance 6, the maximum rule is shown below, and it shows that when there has been decrease in capacitance board 3 activity and in humidity2, there is an increase in other factors' activity and capacitance board 6 activity is maximum.

```
L123456THth
PPPNPPPPPPN
```

7) The below rule shows that when there has been decrease in capacitance board activity 3 and Humidity 2 is decreasing and had led to the maximum activity of capacitance board 6 .

```
L123456THth
NPPNPPPPPPP
```

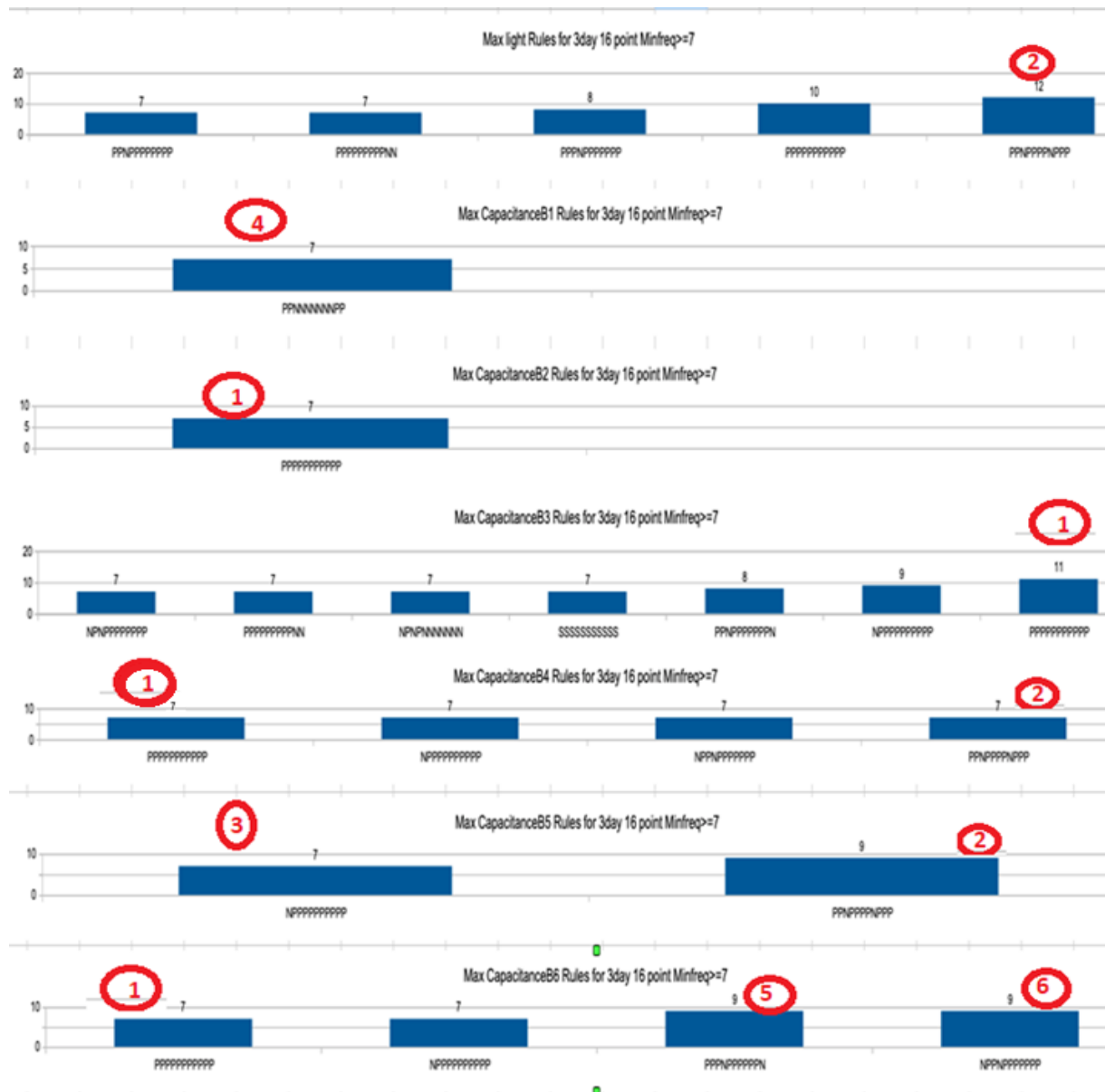The other parts of Figure E.1 do not seem to have such vivid or general rules.

Figure E.1- All-Factors-Relationship rules Frequencies' bar charts for maximum 3-day sliding window covering 16-samples per day when minimum Frequency >=7.

Consider the below chart in Figure E.2, with the focus is finding the maximum activity rules during day when the window size is 7-day and with only 1 sample reviewed each day. The below figure shows that 7-day-1-sample sliding window. It can be seen that, there is not enough frequencies to make frequent rules, because all rules happened only once. The other chart that is for Two-Factors-Relationship's bar-charts in chapter 6 will help to understand what happens in 7-day-1-sample case. The other way would be by decreasing sliding window length and increasing the samples, which is shown in bar-charts in the next section.
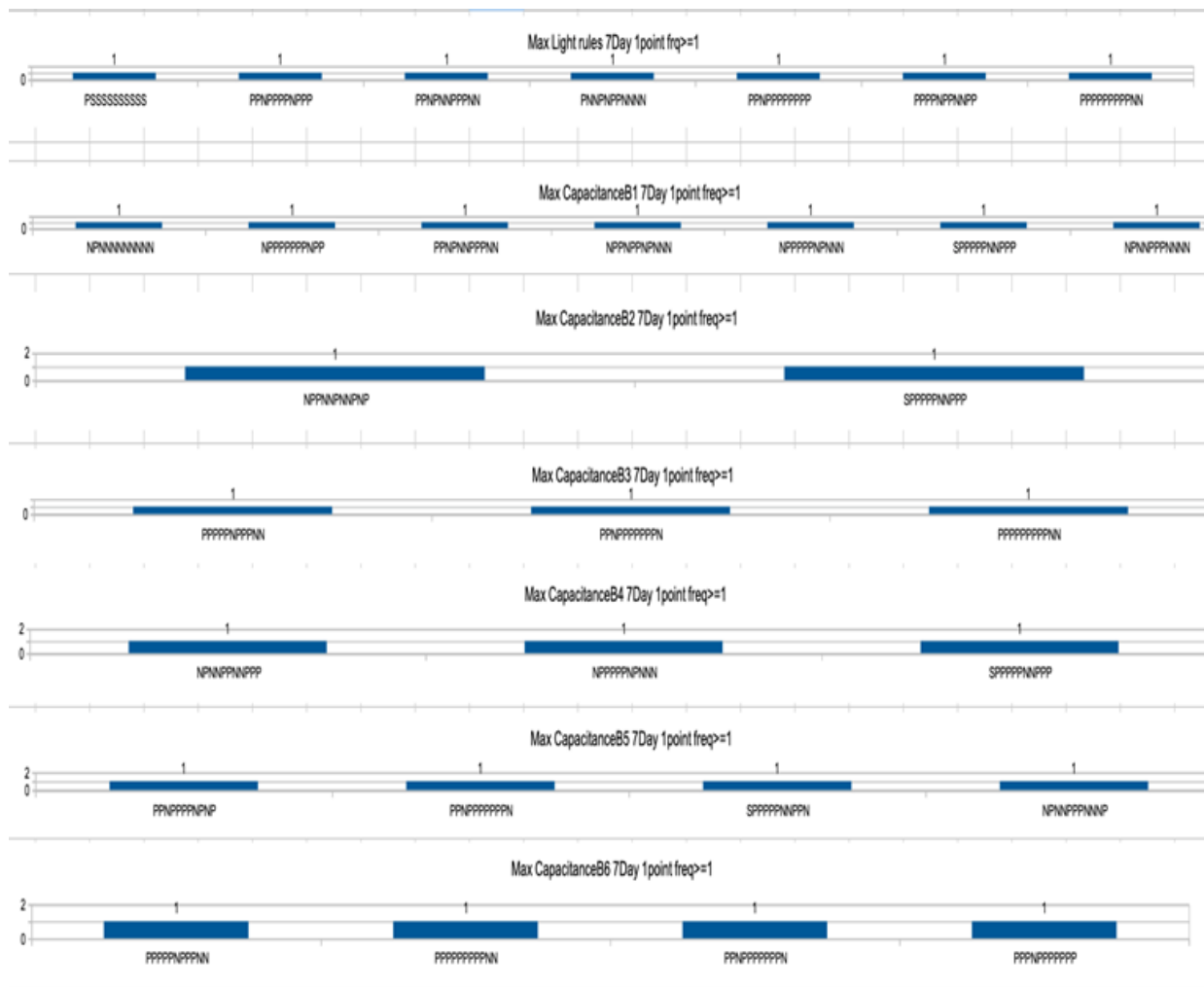

Figure E.2- All-Factors-Relationship rules frequencies' bar charts for maximum-7-day sliding window covering 1-sample in a day when minimum Frequency >=1.

Parts added to software to consider other windows and periods

In this section the new parts have been added to the software, so that it includes 7-day with 16 samples per day sliding window, and finding and focusing on minimum rules of a factor during a sliding window instead of maximum-rules, and night period having 6-samples a night was added. Also, finding how the other factors changed at the same time in the day before.

Consider the following chart in Figure E.3. The focus is finding the maximum activity rules during day when the window size is 7-day-16-samples when minimum Frequency >=7. The paragraph number of each description below is marked by the red circles on the related part of Figure E.3.

1) Consider these capacitance board 2 activity rules. They may show that bees on this board in the brood box, had

L123456THth	L123456THth
PPPPPPPPPP	NPPNNPNNNNN

maximum activities no matter the temperature or light changes. Perhaps this is due to brood and queen being cared for no matter of the environment condition.

2) Also, this capacitance board 6 maximum rule shows that capacitance board 6 would have maximum activity when capacitance board 3 has

L123456THth
PPPNPPPPPP

negative activity changes and all the other factor changes are increasing.

3) Consider these capacitance board 5 maximum activity rules.

L123456THth	L123456THth	L123456THth
PPNPPPPPPP	PPNPPPPNPPP	PPNPPPPPPNP

These it may show that no matter that temperatures are decreasing or increasing, when the

132

capacitance board 2 activity is decreasing the maximum activity of capacitance board 5 has occurred.

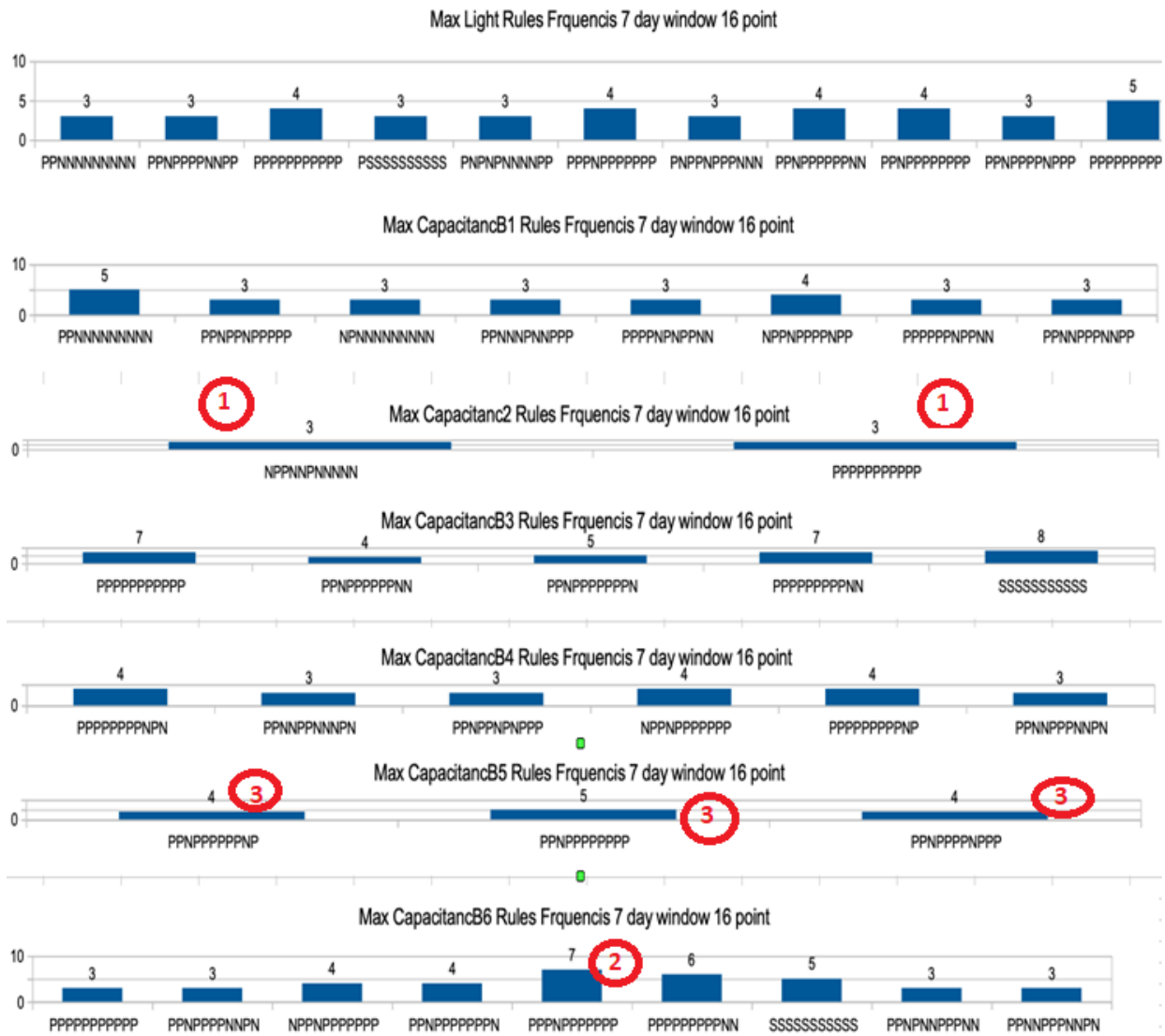The other parts of Figure E.3 do not seem to have such vivid or general rules

Figure E.3- All-Factors-Relationship rules frequencies' bar charts for maximum-7-day sliding window covering 16-samples in day when minimum Frequency >=3.

Charts for obtaining minimum rules

In this section, the software searched for the focused parameter's minimum in sliding window. This will result in minimum (Min) rules. In the chart in Figure E.4, the focus is finding the minimum activity rules during day when the window size is 7-day with 16 samples per day when minimum Frequency >=3. The paragraph number of each description below is marked by the red circles on Figure E.4.

1) Consider the dominant minimum rule of PNNPNPNNNN for capacitance board 1, capacitance board 4 and capacitance board 6. They have the most frequent minimum activities when the temperature and humidity of the boxes have decreased, and the outside light has increased. For other capacitance boards, there is not a vivid dominant rule.

2) Consider Light the dominant rule, which is NNPNNPNNNPP. This shows that even when the light changes amount had been minimum, the capacitance boards 2 and 5 activity changes have increased but other capacitance boards activities has decreased.

3) In Figure E.4, there are two of the most frequent minimum activity rules of the capacitance board 5, PNNPNNNNNNP and NNNPNNNNNNP, that have some similarities between in them. Both rules suggest when the minimum activity of capacitance board 5 has happened, there was an increase in temperature 2, an increase in activity of capacitance board 3, and decrease in other capacitance boards' activities, temperature1 and both humidities. These rules do not dependent on light, since for one rule light is increasing and in another one light has been decreasing.

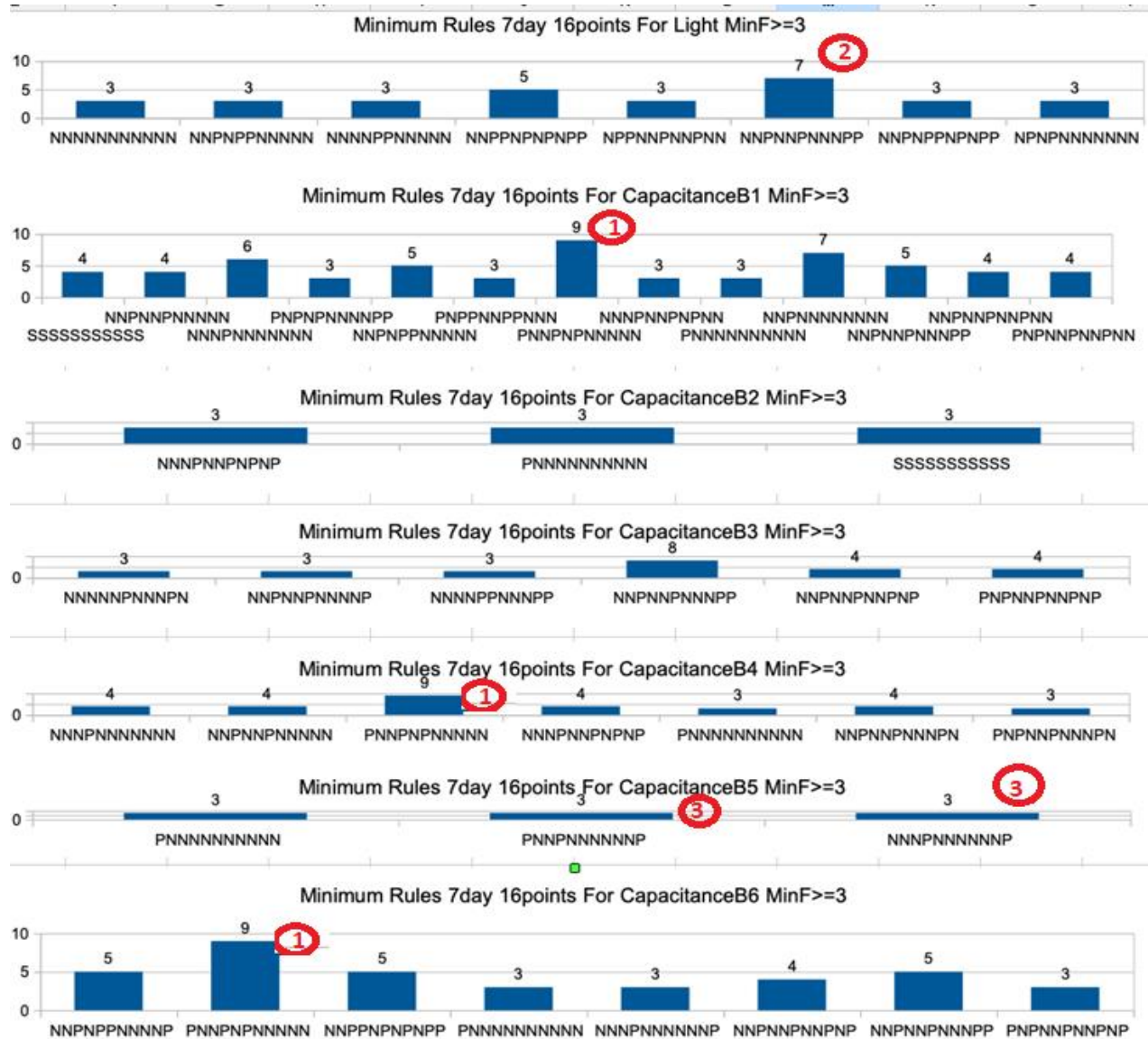The other parts of Figure E.4 do not seem to have such vivid or general rules.

Figure E.4- All-Factors-Relationship rules frequencies' bar charts for minimum-rules 7-day sliding window covering 16-samples per day when minimum Frequency >=3.

Consider Figure E.5, where the focus is finding the minimum activity rules during day when the window size is 3-day with 16 samples per day when minimum Frequency >=6. The paragraph number of each description below is marked by the red circles on the related part of Figure E.5.

1) The dominant minimum rule for light and capacitance board 3 is NNPNNPNNNPP. This shows that when light is minimum and capacitance board 3 activity is minimum, the activity on capacitance board 2 and 5 has increases, also temprature2 and humidity2 have increased. Therefore, it may again be showing an increase in temperature2 and humidity2 causes minimum activity of capacitance board 3.

2) On the other hand, the dominant rule for capacitance board 1 and capacitance board 4 is NNPNNPNNNNN. This shows when the capacitance boards 1 and 4 have minimum activities, most sensors' changes including the temperature and humidity were decreasing in the two hive boxes and only activities of capacitance boards 2 and 5 were increasing.

3) The above mentioned rules differ is in temperature2 and humidity2 value only. As a result, it can show positive temprature2 and humidity2 makes capacitance board 3 minimum activity, and decrease in temprature2 amount and humidity2 can make capacitance boards 1 and 4 have least activity rule. This could be because capacitance boards 1 and 4 were put on the left side of boxes, where they were shaded by trees. This makes their minimum (activity) rule. On the other hand, capacitance board 3 faced more sunshine because (on the right side of the hive) and can become too hot. And so and increase in the temperature2 and humidity2 reduces activity on it. This makes capacitance board 3 minimum rule. It might show a movement of

bees from capacitance board 3 to important hot areas like capacitance board 2 (brood capacitance board) and capacitance board 5 (honey capacitance board), perhaps to help decrease the temperature by fanning. Also, both rules shows that all these three capacitance boards 1, 3 and 4 have the least activity rules when the capacitance board 2 and capacitance board 5 have more activity.

4) Consider capacitance board 2, both its minimum rules (NNNPPPPPPNP and PNNPPPPPPNP) show that the minimum activity of capacitance board 2 happens when temperature 2 decreases and humidity 2 increases, and all other capacitance boards except capacitance board 1 activity increases. This may show that the bees have moved to capacitance boards having honeycomb like capacitance boards 5 and 4, and more honey has been made that can increase the humidity. Also, this rule shows that this bees activity does not rely on light changes, as in one rule it is positive and in another is negative, but in both rules the other factor changes are similar.

5) Consider capacitance board 5. It can be seen that its dominant minimum activity rule, which is PNNPNNNNNNN, shows that it has the least activity when the light and capacitance board 3 activity has increased, but other factor changes has decreased. This may show a previous bee's movement from capacitance board 5 to board 3 when the light has increased, and temperatures and humidity amounts have decreased.

The other parts of Figure E.5 do not seem to have such vivid or general rules.
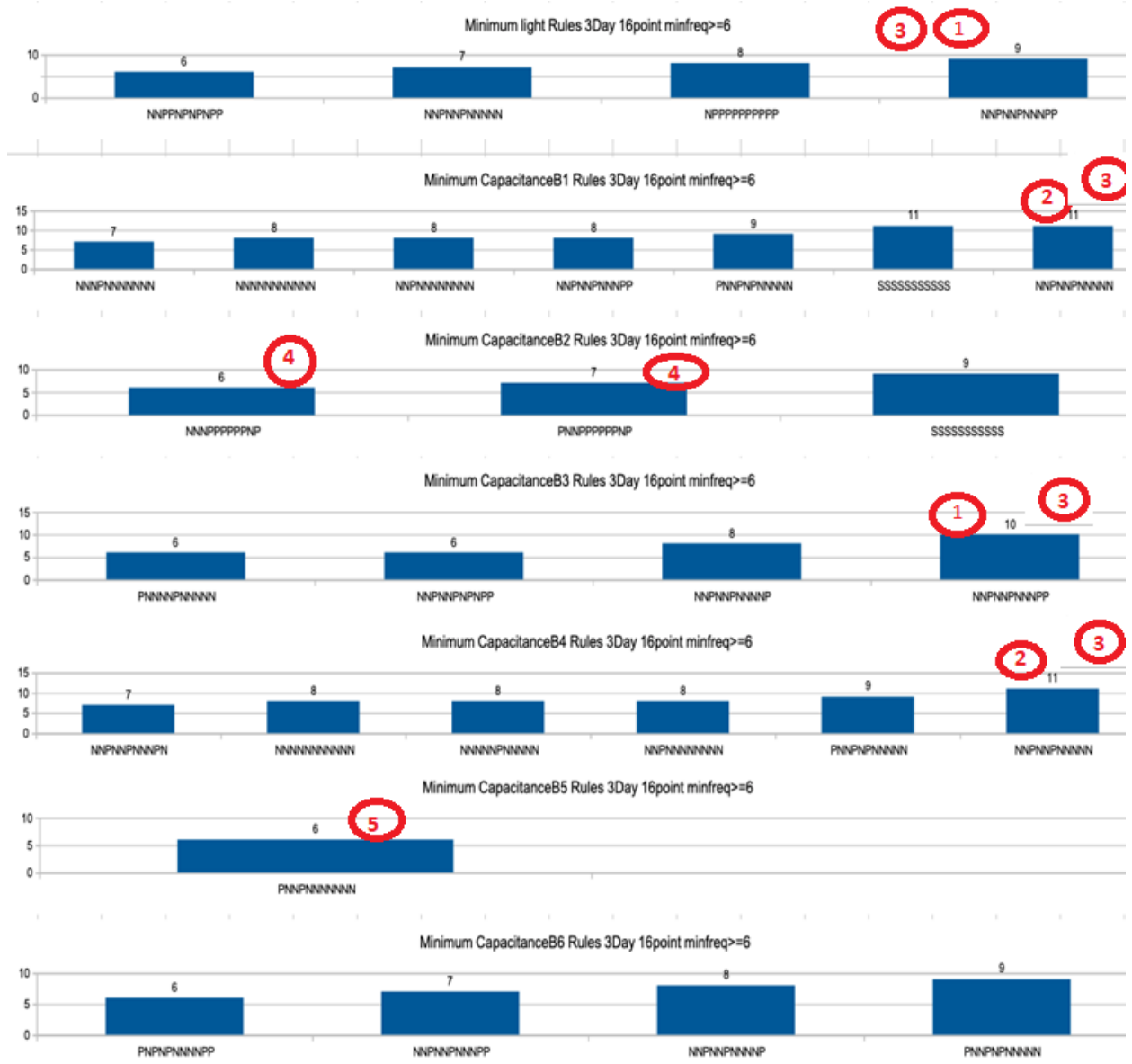
Figure E.5- All-Factors-Relationship rules frequencies' bar charts for minimum-rules 3-day sliding window covering 16-samples per day when minimum Frequency >=6.

Charts for night rules

This section shows data from night studies. It covers 6-samples a night between 00 am to 3 am in a sliding window whether both for it is 3-night or 7-night window.

In Figure E.6, the focus is finding the maximum activity rule during night when the window size is 7-night and 6 samples each night when minimum Frequency >=4. The paragraph number of each description below is marked by the red circles on the related part of Figure E.6.

1) Consider capacitance board 1, board 2, board 4, and board 6. The SPPPPPPPPPP is their dominant maximum activity rule. This show that by increase in the temperature and humidity, these boxes' activity had increased along with other capacitance boards activity increase. It is the second most frequent in capacitance board 5 too.

2) On the other hand, the most frequent maximum activity rule for capacitance board 5 is SPNPPPPNPPP which can show the temprature1 decrease, can decrease activity in capacitance board 2 and increases activity in capacitance board 5.

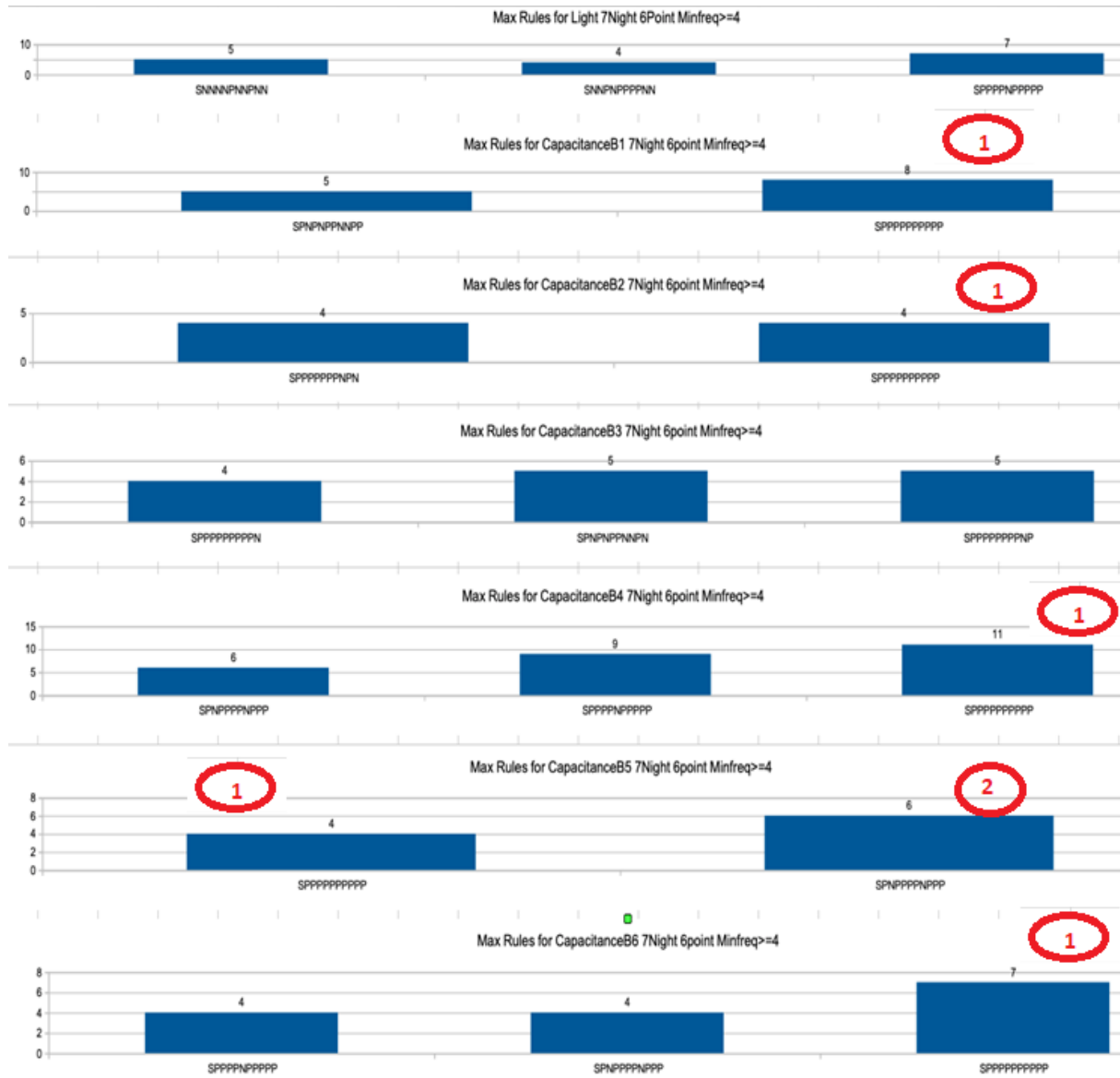The other parts of Figure E.6 do not seem to have such vivid or general rules.

Figure E.6- All-Factors-Relationship rules frequencies' bar charts for maximum rules 7-night sliding window covering 6-samples in night when minimum Frequency >=4.

In Figure E.7 the focus is finding the maximum activity rules during night when the window size is 3-night and 6 samples each night when minimum Frequency >=10. The paragraph number of each description below is marked by the red circles on the related part of Figure E.7.

1) The maximum activity rule for all capacitance boards and light is SPPPPPPPPPPP which can show that with all the boxes' activity increases with the increase in the box's temperature and humidity.

2) The second common rule between capacitance board 1 and 4 is SPPPPNPPPPP. This can show that when the bees have decreased their activity in capacitance board 5 (which has had the most honey at the end), the maximum activity was in capacitance boards 1 and board 4.

The other parts of Figure E.7 do not seem to have such vivid or general rules.

Figure E.7- All-Factors-Relationship rules frequencies' bar charts for maximum rule of 3-night sliding window covering 6-samples in night when minimum Frequency >=10.

In Figure E.8, the focus is finding the minimum activity rules during night when the window size is 7-night and 6 samples each night when minimum Frequency >=3. The paragraph number of each description below is marked by the red circles on the related part of Figure E.8.

1) We see that the dominant minimum frequent rules in capacitance boards 1, 2, 4, and 6, and the second dominant rule for capacitance boards 3 and 5 is SNNNNNNNNNN. This rule can show that when a focused capacitance board has minimum activity (during the sliding window period which is 7-nights here), nearly other bees in other capacitance boards' activity changes is negative. This might happen because of decrease in temperature and humidity of the boxes that is shown in this rule.

2) Capacitance board 5 dominant minimum rule, which is SNPNNNNPNNN, shows that minimum activity of capacitance board 5 happened when there is an increase in the activity of capacitance board 2 and temprature1.

The other parts of Figure E.8 do not seem to have such vivid or general rules.
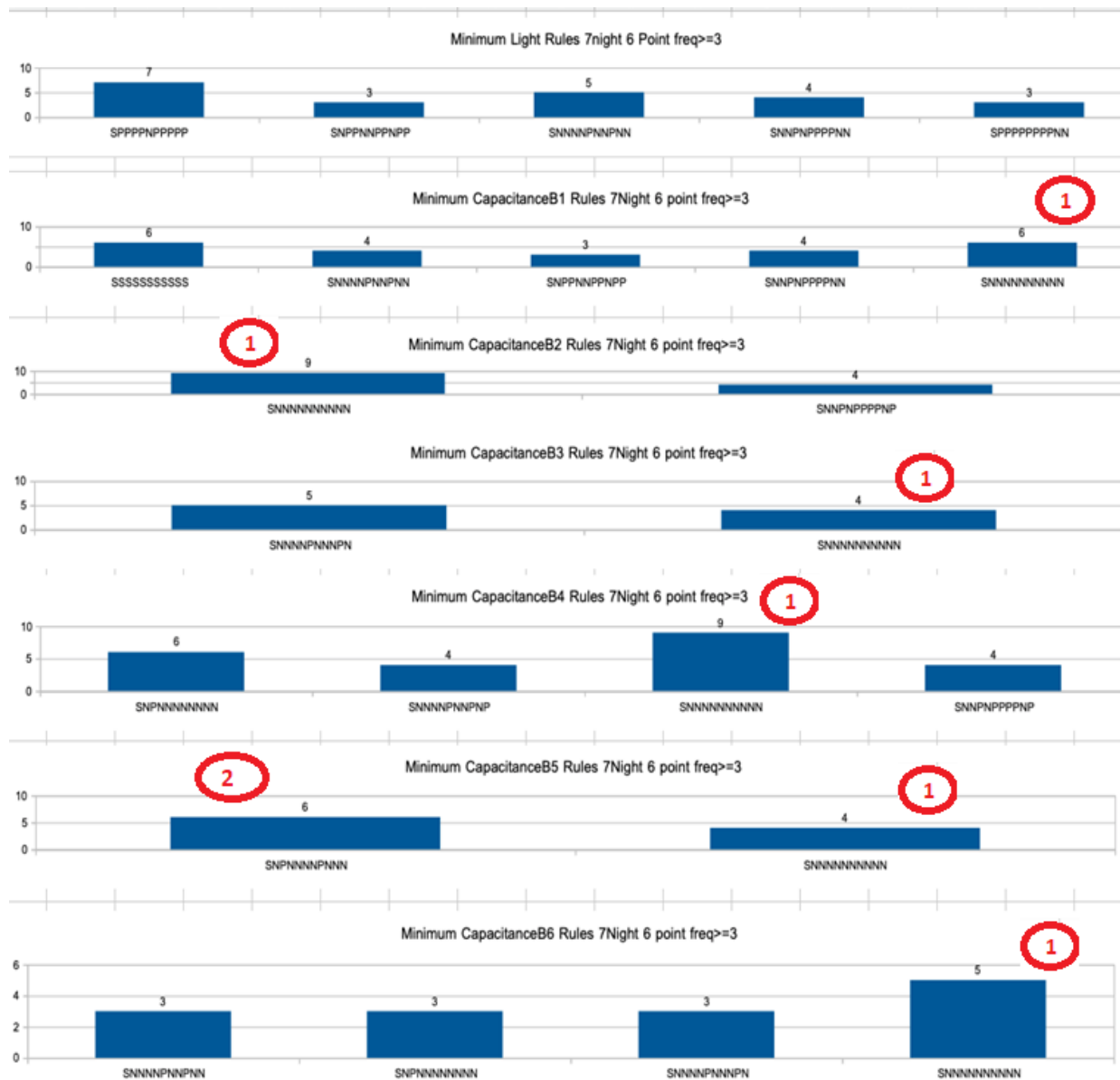
Figure E.8- All-Factors-Relationship rules frequencies' bar charts for minimum rules 7-night sliding window covering 6-samples in night when minimum Frequency >=3.

In Figure E.9, the focus is finding the minimum activity rules during night when the window size is 3-night and 6 samples each night when minimum Frequency >=5. The paragraph number of each description below is marked by the red circles on the related part of Figure E.9.

1) We see that there is a dominant frequent rule in light, capacitance board 1, 2, 3, 4, 5, and 6 which is SNNNNNNNNNN. This rule can show that when a focused capacitance board has minimum activity (during the sliding window period which is 3-nights here), the activity of bees on most other capacitance boards is negative too. Moreover, it shows that the temperature and humidity in these capacitance boards have decreased too. In other words, the figures above show the most frequent minimum rules among all sensors when the light and capacitance boards are focused, and are minimum in their sliding window is SNNNNNNNNNN. It shows that in this most frequent rule the light has been constant and the focused sensor along with other sensors had decreasing values.

2) A common second minimum rules between capacitance board 1, 3, 4, 5, and 6, and light is SNPNNNNNNNN.

The other parts of Figure E.9 do not seem to have such vivid or general rules.
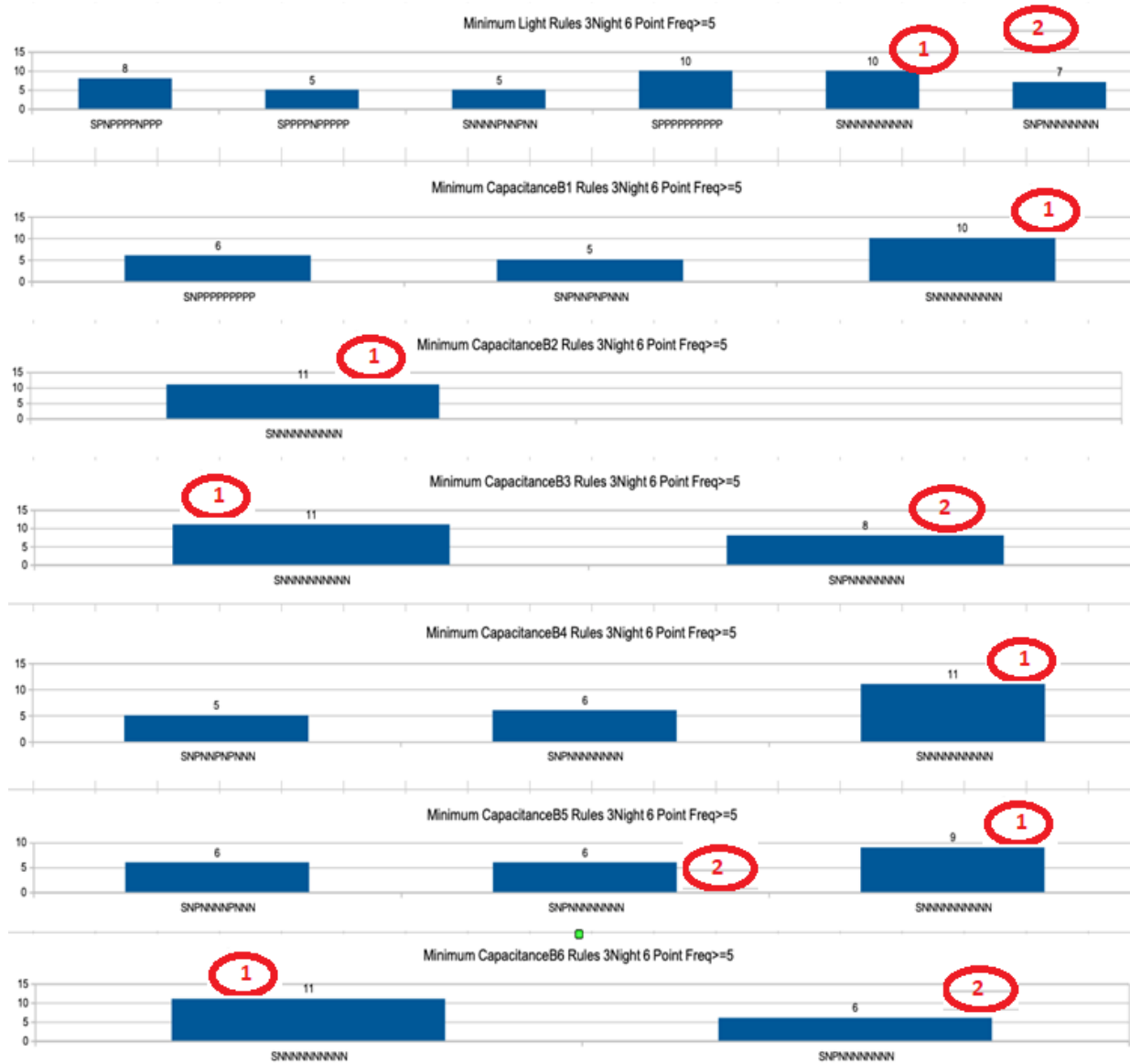
Figure E.9- All-Factors-Relationship rules frequencies' bar charts for minimum rules 3-night sliding window covering 6-samples in night when minimum Frequency >=5.