Exploring the Effects of Cell Structure on Plant Morphology Using Time-Series Microscopy and Computer Simulations of Moss Growth

A Major Qualifying Project
Submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE
In Partial Fulfillment of the Requirements for the
Degree of Bachelor
of Science
By:

_____
Max Levine

Date:
May, 2015

Approved:

_____
Professor Luis Vidali
Worcester Polytechnic Institute

Abstract

As the global population increases, it becomes increasingly important that we find ways to improve our agricultural sustainability. In order to provide the necessary improvements to keep up with agricultural demand, we must advance our understanding of how plants grow and thrive. In this work, I designed methods to assist in the process of ascertaining this knowledge from the model plant, the moss *Physcomitrella patens*. I developed a new growth plate which allows the moss plants to be kept alive, and imaged under the light microscope, for an extended period of time. The use of this plate will help improve our understanding, through time-series microscopy of live plants, of the relationship between cell growth and plant morphology. I also evaluated the effectiveness of a growth assay that is used to characterize differences in growth patterns. This was done by exploring the parameter space of a moss growth computer simulation. This allowed me to determine how changes in cell structure affect the plant morphology metrics of the assay. Additionally, a K-nearest neighbors model was developed to help predict, using morphological metrics, changes in cell structure of possible mutant plants. The improved assay and model will be used to investigate how genetic changes that affect cell growth alter plant morphology.

Acknowledgements

Table of Contents

Table of Figures

Table of Appendices

1 Introduction

1.1   Importance of Understanding Plant Growth

In 2013, the Food and Agriculture Organization of the United Nations produced a study

estimating that, to sustain the rising human population, the world will need agricultural products

to be grown in greater quantities and in a more efficient manner. They stated that, by 2050, 60%

more agricultural products would need to be grown compared to 2005-2007 amounts (FAO,

2013). Further, in addition to more farms, the world will need to improve plant growth

efficiency. Therefore, to accomplish this, we must better understand how plants grow so we can

determine how to best increase growth amount and growth speed. By studying model plants, we

learn about the different systems in plants and then look to apply this knowledge to agricultural

plants (Flavell, 2009).

1.2   Model Organism *Physcomitrella patens*

One particular plant used to study plants and plant growth is the model organism

Physcomitrella patens ("*P. patens*") because plants similar to mosses are some of the earliest

land plant life (Schaefer and Zrÿd, 2001; Knight and Perroud, 2010). An image of *P. patens*

mature gametophores can be seen in Figure 1.

In the wild, *P. patens* plants begin as spores yet, in the laboratory they can be cultured

from single cells called protoplasts, which means that the growth conditions are artificial to an

extent. The plants start as single cells, which are spherical(either spores or protoplasts). When

protoplasts are regenerated, they are cultured in mannitol, which stabilizes their osmotic

pressure. Under these conditions, the protoplasts form strings of 1-5 "bulbs", which are close to

spherical cells. Once the bulbs are placed on growing media, branches then grow from these bulbs, the branches curve and other branches then grow from the original branches (see Figure 2). These simple plants make moss an excellent species for studying plant growth and development since many plant growth genes are highly conserved in more complex plants (Prigge and Bezanilla, 2010). Furthermore, *P. patens* plants are used often in studies because they are easy to culture in vitro. *P. patens* is also highly efficient at gene targeting as this moss allows easy identification of mutant phenotypes since the plant is haploid, and its genome sequence is known (Knight and Perroud, 2010). This moss has already been highly studied in several areas in plant biology including its polarized growth and genetic analysis (Schaefer and Zrÿd, 2001; Vidali and Bezanilla, 2012).

Figure 1: Image of Physcomitrella patens

Areas that have been less studied are quantifying moss growth, including 3D growth, and structural changes in mutations, which are further developed in animal model systems (Deal et

al., 2016). By increasing our knowledge of these areas, we will be able to better measure plant changes and create tools utilizing a quantitative approach that will allow predictions of plant structures. Creating a prediction tool will enable faster screening of the structure of mutant plant lines to provide quick analysis of how mutations affect growth.
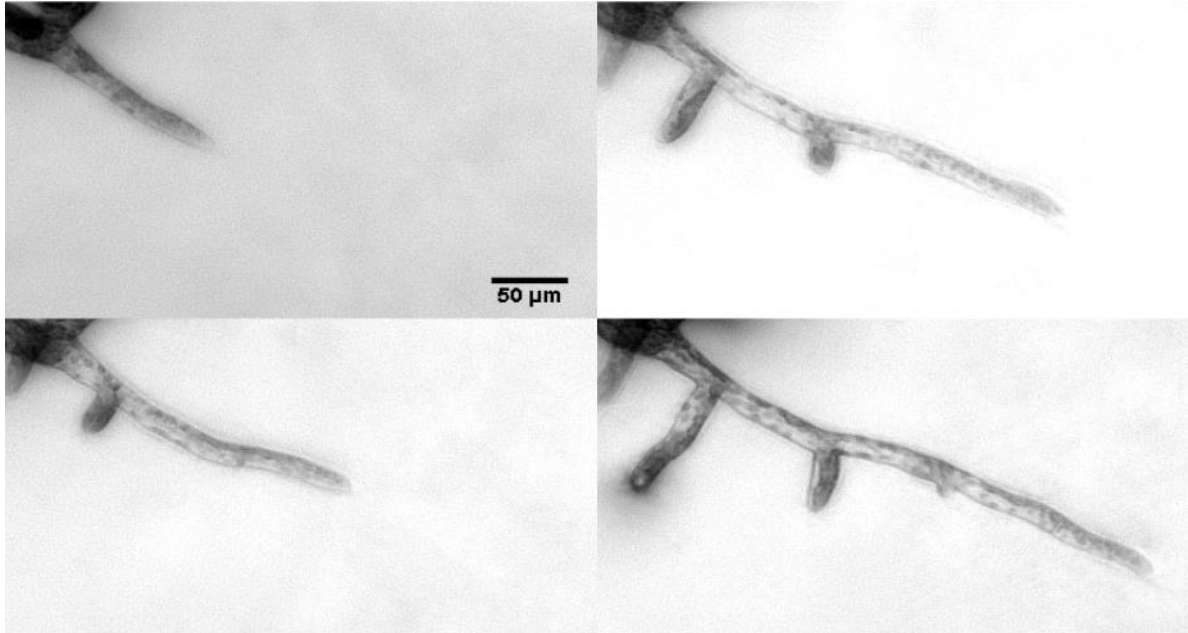
Figure 2: Images showing P. patens branching

1.3   Plant Morphology Analysis

Quantifying the morphology of a plant improves our capacity to analyze the growth of concerted plant cells for comparative studies. In the Vidali Laboratory at WPI, a methodology was developed to quantify the morphology of images of *P. patens* plants (Bibeau et al., 2014). An ImageJ macro, which was developed in the laboratory, analyzes the morphology of the plants using area, convex hull area, perimeter, and convex hull perimeter. These values quantify the plant using the metrics: area, eccentricity, solidity, circularity. Eccentricity is a measure of how elliptical the entire moss plant is. Solidity is a measure of area vs. convex hull and gives an idea

of the density of a plant. Circularity is a measure of area of the moss vs. the area of the circle that has the same perimeter as the moss. A visualization of circularity, eccentricity, and solidity is shown in Figure 3.

Hundreds of plants were analyzed with this 4-metric macro and statistical methods were used to test differences in morphology between temperature sensitive plant mutants (Ding et al., 2014). The methodology was used on each plant at the same time step, thus analyzing solely morphology, not growth over time.

| Circularity | 1 | .9153 | .8760 | .6142 |

| Eccentricity | 0 | .661 | .8660 | .9682 |

| Solidity | 1 | .681 | .578 | .488 |
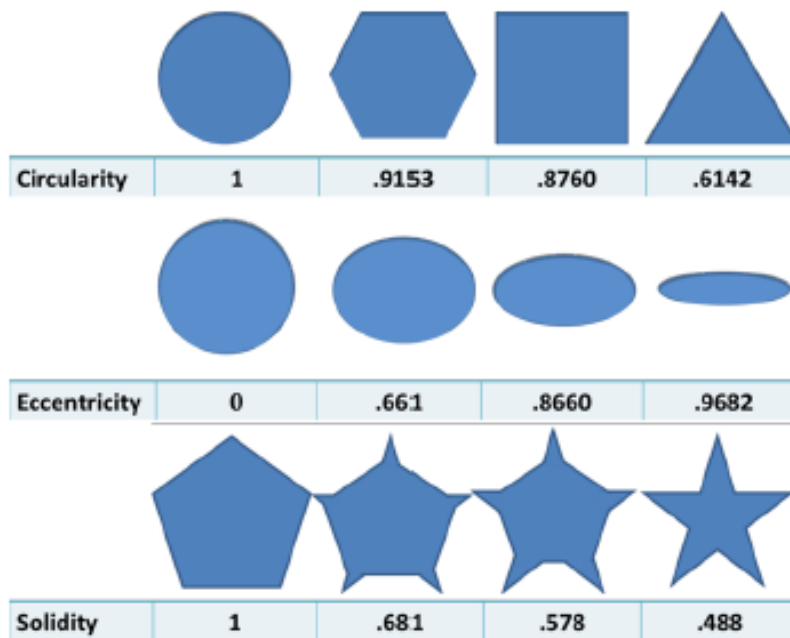
Figure 3: Visualization of circularity, eccentricity, and solidity (Lemoi and Tüzel, 2012)

Fractal dimension was introduced as a new metric in addition to the four previously mentioned metrics in a recent WPI Major Qualifying Project (Lemoi and Tüzel, 2012). Fractal dimension is a measure of how space-filling an object is, but its calculation is not trivial for non-regular shapes. To calculate fractal dimension for an arbitrary image, box counting method is used (Lemoi and Tüzel, 2012). The concept of this method is to count the number of boxes(N)

that the image fills depending on the size of the boxes (s) and the fractal dimension is the ratio of how those numbers scale. This method is illustrated in Figure 4.
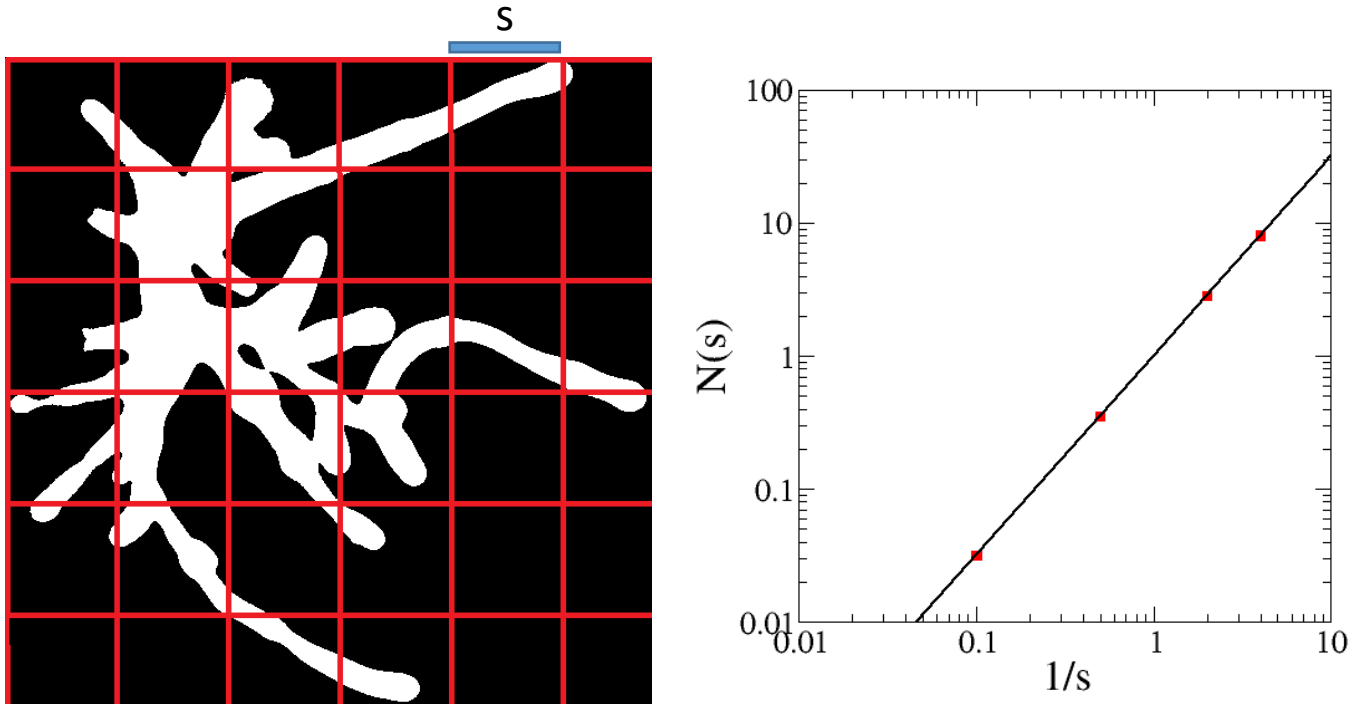
Figure 4: Box counting method for determining fractal dimension. Fractal dimension is the slope of the graph of log(N) vs log(1/s), where N is the number of boxes and s is the side length of the boxes.

1.4   3-D and 4-D Imaging of Plants

To analyze plant structure the plants must first be imaged. Two dimensional images can easily be captured with many different microscopy methods and current morphology assays use these type of images (Bibeau et al., 2014; Deal et al., 2016). Three dimensional images can be reconstructed using slices of the sample at set intervals with a technique called tomography to also capture depth information (Schmid et al., 2010). The three dimensional reconstructions can help resolve overlaps from the projection of the specimen features on themselves that occur in two dimensional images (Truernit et al., 2008). Light Microscopy and Fluorescence Microscopy

are the most common acquisition types used with tomography (Eckardt, 2008; Domozych, 2012). Light Microscopy, where images are generated by light that is transmitted through the sample, is often used because it is inexpensive and readily available; but the images acquired have a large amount of background noise making three dimensional reconstructions harder. With Fluorescence Microscopy, fluorescent probes can be used to highlight sites of interest in the sample and there is much less background noise compared to light microscopy (ThermoFisher, 2009). In addition for three dimensional reconstructions, techniques such as deconvolution, confocal microscopy, and the ApoTome, can be used to remove some of the out of focus light to sharpen the quality of the 3-D images (McNally et al., 1999).

Four dimensional images are created by taking three dimensional images at set time intervals. Some issues arise when time is factored in: the sample must be kept alive and in controlled conditions over the time period, the light from imaging can cause harm from photodamage, and the fluorescent dyes can affect the specimen (Murphy and Davidson, 2012). Due to these challenges, four dimensional images of plants growing have not been explored in depth.

1.5   Simulated Plant Growth

Modeling plant life is a relatively new field for using artificial intelligence to generate plants and plant behaviors. To virtually grow the plants, recursive algorithms are utilized. The most common technique is L-system, which involves string rewriting (Steinberg and Sikora, 1999). L-system starts with an initial node represented by a string a letters and is recursively rewritten based on a list of rules. These systems can model many plant types relatively well, including their root structure (Stefan and Claude, 2007).

A concern of using L-system is that the lists of rules can get increasingly more complicated, making the model hard to implement. Another implementation of the L-sytem method, which is not as constricted as simple L-system, is using a procedural algorithm, where conditions, such as variables contained in different parts of the plant, are checked to determine in the next corresponding step, to grow the plants. This is different from simple L-system since simple L-system is generally deterministic whereas in procedural algorithms decisions are often selected from ranges or probabilities. Procedural systems also often have variables which are not directly represented in the plant structure. Experimental data can be put into this algorithm to try to replicate plants that look like those in the experiments (Lemoi and Tüzel, 2012).

The Tüzel group at WPI Physics Department developed a *P. patens* growth simulator, which is based on experimental data to simulate populations of moss plants using procedural generation (Vandal et al., unpublished). The simulator uses inputs including cell length, cell width, curving angle, and branching angle to grow the plants, which have been tuned to what is observed in the lab. The program first generates a set of bulbs then adds branches to the bulbs. These branches are lengthened and then new branches are added on to original branches. In Figure 5, you can see that the simulated plants (bottom) look similar to experimental plants (top). Based on the metrics (area, solidity, circularity, and eccentricity), explained in Section 1.3, Vandal et al. have shown that the model very closely matches the experimental data.
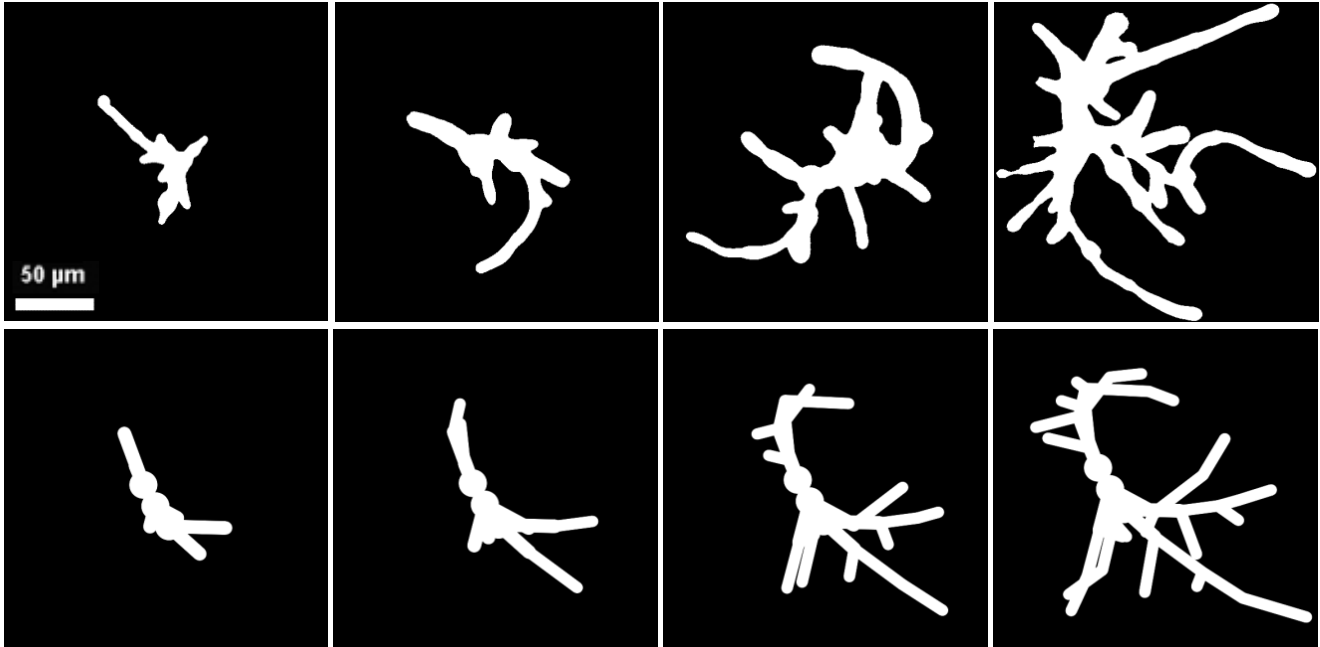
Figure 5: Comparison of experimental (top) and simulated (bottom) moss plants.

1.6 Purpose of Project

The first goal of this project is to develop an easy-to-use growth assay for analyzing the growth of plant cells over time. To do this, I will develop a methodology for obtaining four-dimensional images. When fully implemented, this assay should be able to compare differences in growth between plant species and mutants, taking into account: total volume, average cell volume, cell number, cell shape, change in volume over time, change in cell number over time, and plant shape. To ensure the ability to measure growth over time, it is important that this assay is not invasive to the plant and does not affect the plant growth/viability.

The second purpose of this project is to test how well the automated morphological assay, mentioned in Section 1.3, is at differentiating changes in plant structure. Using an existing moss growth simulator, developed by Vandal et al., I will explore how well different quantitative

metrics describe different plants (Vandal et al., unpublished). I will then try to predict the structure of experimental plants using just quantitative metrics that can be gathered automatically so that we can easily screen mutations.

2 Materials and Methods

2.1 Culturing *Physcomitrella patens*

In order to learn about the growth of *P. patens,* we need to be able to culture the moss quickly and effectively. The cells of the moss are totipotent, which means that the whole plant can grow from just a single cell. This means that we can explore *P. patens* growth from single cell onward. Therefore, we need to be able to collect these single cells.

To culture single cells, we first break down the cell wall of 7-day old moss using driselase. We preform several washes of the cells, which are now called protoplasts, so that the driselase does not kill cells. The protoplasts then need to recover their cell walls so that they continue to grow. To achieve this, we then plate them on top of cellophane on PRMB plates for 4 days. The plants are then ready to image. For this project, I only looked at the wild type for *P. patens,* which is plant cell line 576 in our lab.

2.2 Staining the Moss

In the interest of creating solid three dimensional reconstructions of the plants, I needed to find a stain that would not affect the growth of the plants and would also highlight the shape of the plant. The first dye that I looked at is propidium iodide, which is often used in viewing the cell wall of plants, since it binds to structural pectin (Rounds et al., 2011). Propidium iodide, which generally fluoresces red, has the added benefit of staining dead cells bright yellow so you can easily distinguish them and it fluoresces under green light, which is not detected by the moss. However, after testing, I determined that the dye affected the moss and prevented it from growing (see Figure 6). I suspected that this was due to photodamage. To prove this point, I

diluted the dye and found that, when diluted down to a level where phenotype is not affected, the
dye could not be seen.
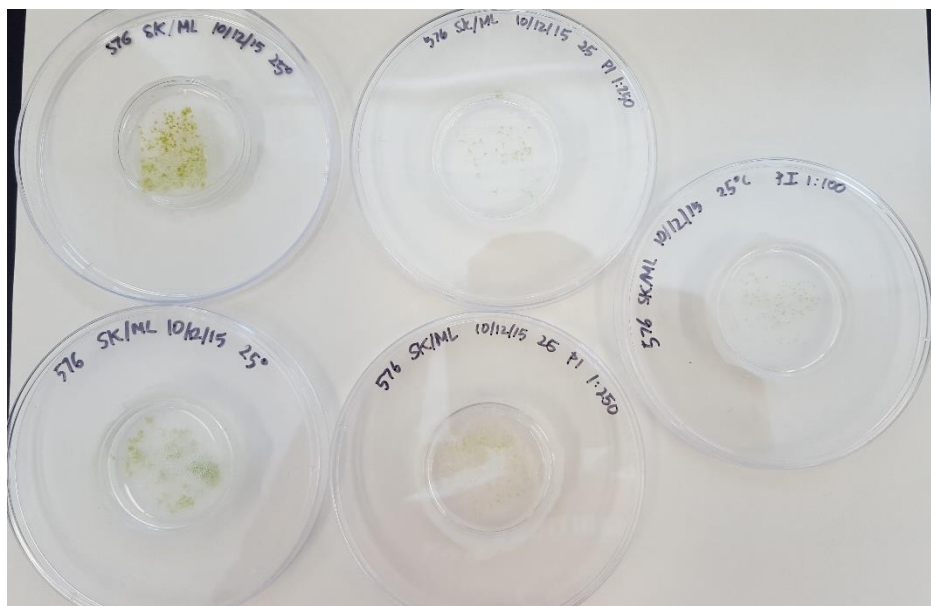
Figure 6: Reduced growth with prodidium iodide stain. Increasing amounts of dye from left to right.

My second choice of dye was calcofluor white, which binds strongly to cellulose, which
means we can use it to stain the cell walls of the moss. I found, if diluted to 1/200 of a 1mM
stock or more, the dye has little effect on growth and viability of the moss. The only issue with
using calcofluor is that it needs to be illuminated with ultraviolet light. Ultraviolet light can be
detrimental to moss, killing it or causing mutations. I had to be careful that I did not use too high
intensity light and used short exposures of 100ms so that I did not overexpose the moss. I ended
up using a final concentration of 5 µM calcofluor in media to minimize negative effects of the
dye while also being able to capture the color under a microscope.

2.3 Keeping the Plants Alive During Imaging

Another big issue with trying to image plants over a longer period of time is providing conditions for the plant to continue to grow during imaging, while still being able to image it. My first attempt was to use a large glass slide that had a square bounded off with double sided foam tape. I then filled the square with medium, transferred the protoplasts on top, and then covered it with a large cover slip. Although I was able to easily image the slide, it would dry out quickly and the plants stopped growing. I obtained similar results when I used a small thin plate.

My current setup involves making a plate that has thick medium around the outside and thin in the middle so that I can image through the center. This formation, shown in Figure 7, prevents the moss from drying out because the moisture in the outside ring replenishes the moisture in the inner area. I only image plants that are close to the outer ring to minimize moisture loss.
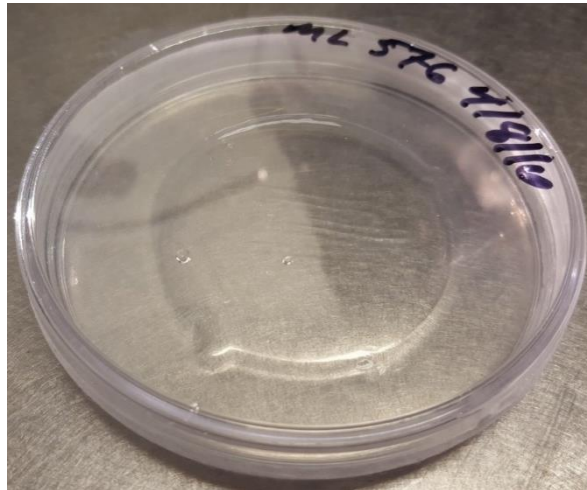
Figure 7: Long term imaging plate

To create this special plate, I first take a 90mm plate and put a 60mm plate inside of it with a weight to hold the inner plate down. Sixteen mL of ppNH$_4$ media, with 5 µM calcofluor, is

12

then poured into the space in between the two plates. Once the media has set, I remove the inner plate leaving an empty cavity in the center. The cavity is then filled with four mL of ppNH$_4$ media, with 5 µM calcofluor, and, once set, the plate is ready for the moss.

To plate our 4 day old protoplasts, which are on cellophane, I cut a piece of cellophane that is smaller than the inner plate area. I then flip the cellophane piece over onto the inner part of the plate so the protoplasts are directly contacting the media. After a minute, the cellophane is peeled off the media so that a majority of the protoplasts are left on the plate. The plants are then ready for extended period imaging.

Another challenge with long term imaging is keeping the plants illuminated so that they continue to grow while under the microscope. Our experiments showed that the bright field lamp on the microscope was not sufficient enough for the plants to grow. The incubator that we use to culture our moss illuminates the plants with a little over 100 micro mol/cm^2; so I need to achieve this on the microscope. To do this, I use an external lamp that illuminates the plate from the side with a little over 100 micro mol/cm^2. Our set up can be seen in Figure 8.

During imaging I use an autofocus algorithm to ensure that the plants are always in focus over the course of acquisition and try to only image plants near the outter ring to decrease the chance of the plants drying out. I also set up the exposure of the ultraviolet light, used to fluorese

the calcofluor, to be very low (100ms) and only on while the camera is taking a picture to prevent damage to the plants.
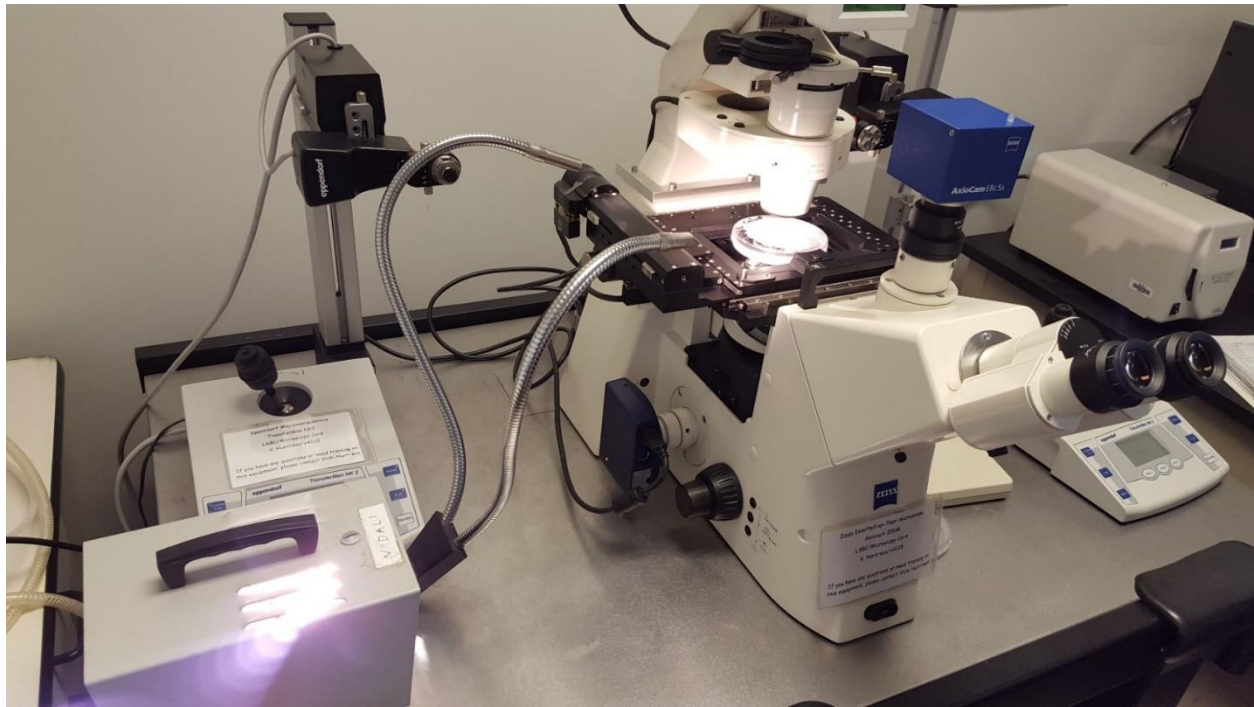
Figure 8: Microscope setup with external lamp for long term imaging.

2.4   Generating a Database of Simulated Plants

In order to analyze the quality of our quantitative metrics for differentiating mutations in plant growth, I first needed a large set of plants with different types of mutations. The most logical solution was to generate simulated plants for which we know all of the parameters. To do this, I used the moss growth simulator developed by the WPI Physics department, as seen in Figure 5 and noted in Section 1.5. Although the default parameters for the simulator are experimentally obtained data, I can change these parameters to create mutants. The four main parameters, relating to plant morphology, that I explored are cell length, cell width, curving angle, and branching angle.

I created a database of 2,401 simulated plant populations of 100 plants by using all combinations of 7 values for cell length (20, 40, 60, 80, 100, 120, and 140 percent of experimental), 7 values for cell width (40, 60, 80, 100, 120, 140, and 160 percent of experimental), 7 values for curving angle (1, 15, 30, 45, 60, 75, and 90 degrees), and 7 values for branching angle (0, 15, 30, 45, 60, 75, and 90 degrees). The datasets were created using a batch file (see Appendix 1) that would run the growth simulator, which is written in Java, with the different combinations of parameters.

The simulator also has parameters for length, curving angle, and branching angle standard deviations. I kept the curving angle and branching angle standard deviations the same as for experimental, but I changed the length standard deviation to be a set percentage of around 31% of the length parameter, determined from the experimental data. This was done in order to keep the variation of the length of the plant cells within a plant realistic.

For the purpose of the research I only look at 3 day old simulated plants, since the growth simulator has been shown to match the experimental plants, based on the morphological assay, for 3 day old plants.

3 Results

3.1 Extended Time-Series Imaging

Using the methodology outlined in Sections 2.1-2.3, I was successfully able to capture time series images of moss growth. In addition, by using a microscope equipped with an automated stage and focus, I was able to image multiple plants at a time and take z-stacks, which could be used to create three dimensional reconstructions of the plants.

Figure 9 shows images from a representative time-series; these images were collected during a 48 hour acquisition span, with images being taken every thirty minutes. The plants growing under the microscope seemed to grow faster than plants left in the incubator, although establishing if this is a consistent observation will require additional experiments and measurements, to be able to accurately compare the size of plants on the microscope and plants from the incubator. A possible explanation for this observation is that the plants under the microscope have increased light exposure, in particular because they do not experience the light/dark cycle of the plants in the incubator.

Importantly, it does not appear that the moss plants suffer from any negative effects from our imaging methodology. This is important because the plants are under constant exposure to calcofluor and exposed at regular intervals to near UV radiation used for imaging. As seen in Figure 9, the calcofluor is a good marker for the internal area of the plants, since it clearly highlights the cell walls. It is even possible to visualize the cell wall forming during polarized growth and during cell division. Unfortunately, it appears that the dye becomes less bright over time, which can make the plants harder to distinguish from the background at later time points.

This fading is most likely caused by photo bleaching of the calcofluor and its slow exchange rate, since all of the media surrounding the plant contains the dye.
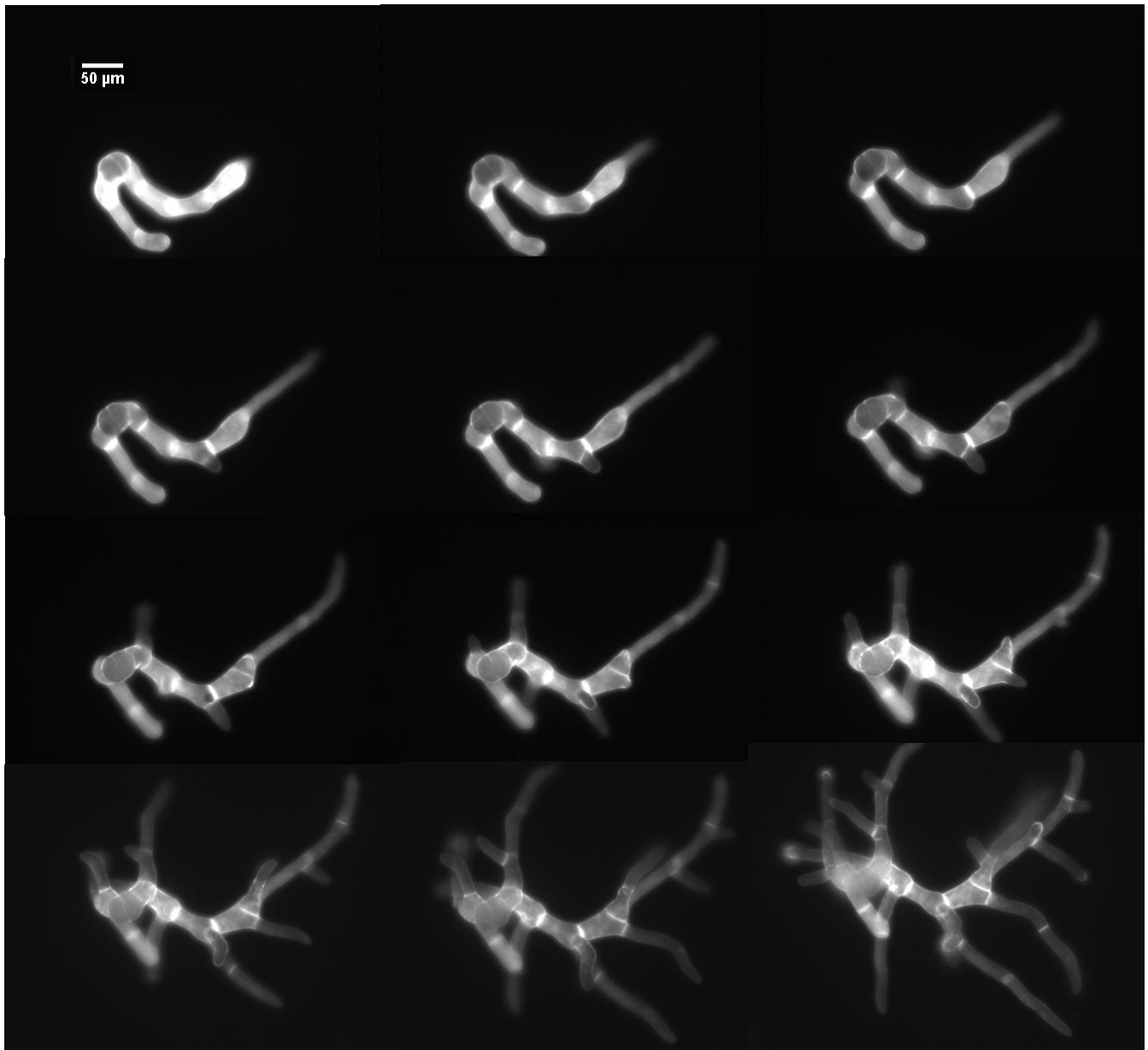
Figure 9: Selected frames from 48 hour time series of moss growth.

This methodology that I developed is important because it allows individual plant samples to be studied for an extended period of time while they grow. This is a significant improvement

because previously, in the laboratory only populations of plants could be studied, by taking out the cultures from the incubator, imaging them, then putting them back in the incubator.

3.2   Three Dimensional Reconstructions

Using optical sectioning (z-stacks) it is possible to collect volume and depth information that is not available with regular two dimensional imaging. As part of this project, I was able to use the Apotome function of the epifluorescence microscope, combined with three dimensional rendering software to create three dimensional reconstructions of moss protonemata. A sample of one of these reconstructions is seen in Figure 10. This new information includes plant growth
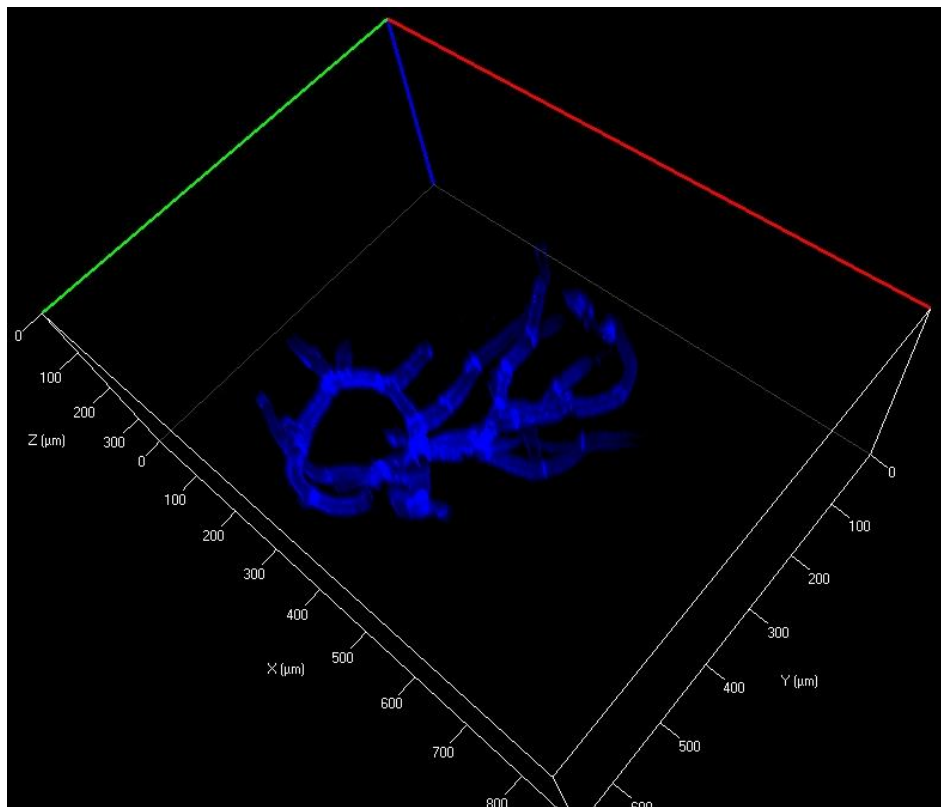
Figure 10: Three dimensional reconstruction of moss plant using calcofluor and z-stacks. The images were collected using a Zeiss ApoTome microscope attachment, which allows for only in-focused light to be captured. The reconstruction was then made using the Zeiss built-in three dimensional rendering software.

behavior towards or away from the light source and how the plant interacts with itself during growth. Three dimensional reconstructions are important because they collect information that may be lost in overlapping that can occur in two dimensional projections. Unfortunately, I did not have enough time to optimize imaging conditions to obtain reliable three dimensional images over time. In addition, the Apotome function was not present in the microscope with the automatic stage, although deconvolution algorithms may be useful to remove out of focus light as an alternative to the Apotome.

3.3   Morphology Assay Analysis

Using the large database created with the conditions outlined in Section 2.4, I tested the effectiveness of the morphological assay described in Section 1.3. My initial goal was to evaluate the metrics from the growth assay; hence, I applied these metrics on the simulated plants to determine if they are good at differentiating moss populations with different growth parameters. If the metrics are adequate at differentiating plant populations resulting from changes in parameters, then, when plotted by metrics, we should see plants from a population of parameters clustered together and away from other populations. Figure 11 shows this kind of plot of area vs. solidity for different changes in the parameter for cell length (Code to generate this is seen in Appendix 3). To demonstrate how well separated the clusters are, I superimposed the 95% confidence regions for each population on Figure 11 (seen in Appendix 2). The figure in Appendix 2 shows that the 95% confidence regions highly overlap which means the populations are not statistically separated and therefore we cannot detect differences of any less than a large percentage for cell length.
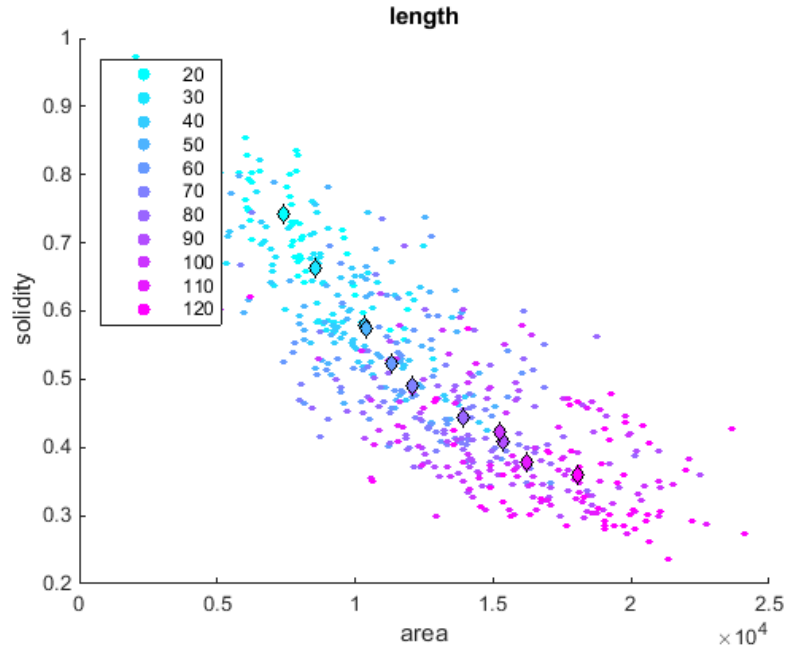
Figure 11: Populations with varied cell length plotted against evaluation metrics. The large diamonds show the means of the plant populations.

To be useful, metrics should show a significant trend for changes in a parameter; hence, I wanted to explore the existence of these trends in the available metrics. To help visualize the existence of these trends, I plotted box graphs for the metrics for different values of the parameters. One of the box graphs for cell length parameter is shown in Figure 12. As a first approximation, I use these plots to determine which metrics are not useful (do not show a trend with changes in the parameter) and to more accurately see how big a difference we can detect for the metrics. These box graphs also give us a better idea of the amount of variation there is within a population. From the box plot in Figure 12, we learn that eccentricity does not show a trend, fractal dimension has a weak trend, and that the other three metrics (area, solidity, and circularity) have stonger trends.
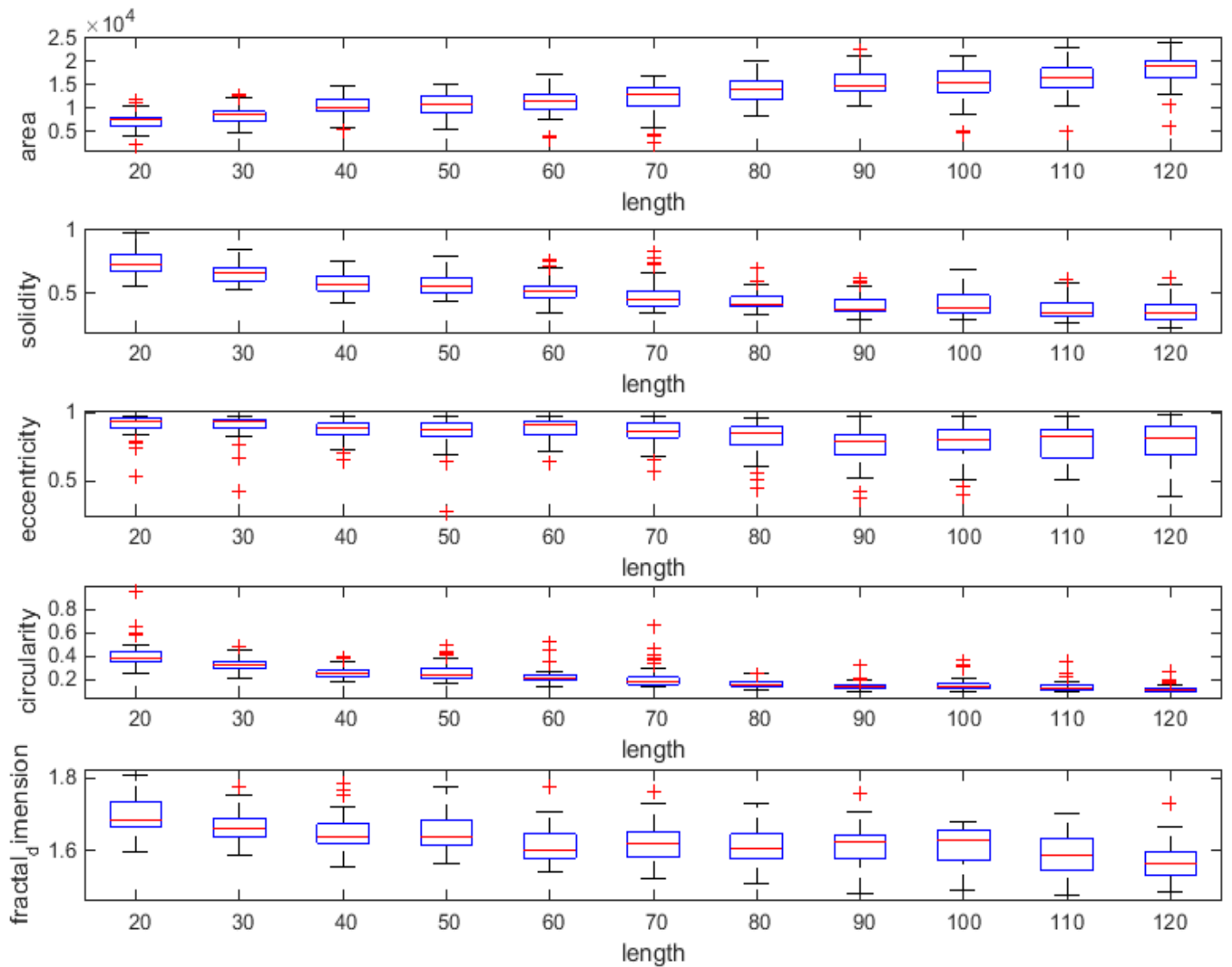
Figure 12: Box chart to visualize trends in metrics against on parameter.

The next way that I tested the metrics was to look at how they varied with changes in two parameters at time. To do this, I generated two dimensional heat maps of the metrics' mean values for populations that vary based on two different parameters. These maps for cell length and width can be seen in Figure 13. The goal was to be able to visualize different trends for each metric in the set when changing two parameters at a time. For example, in Figure 13, we see that

eccentricity does not show any trend for day 3 plants with varying cell length and width, which suggests that this metric may not be useful. I also noticed that solidity, circularity, and fractal dimension all show the same general trend for this scenario, which probably means we are collecting redundant information.

As a first approximation to statistically test if the metrics can differentiate populations of two parameter variation, I generated T-test heat maps. In these maps, I compare the metric values for each population against the wild type to see if they are statistically different, with at least 95% confidence. One of these sets of t-test heat maps can be see in Figure 14. In this figure, I color the wild type population with green and highlight the populations that are not statistically different from the wild type in yellow. A good metric would have no yellow because that means that the wild type is statistically different from all the other tested populations. So looking at Figure 14, we see that eccentricity is not useful in differentiating cell length and width variations of day 3 moss. We also learn from this figure that, even if the heatmap shows a trend, that does not mean that we can differentiate every population when varying two parameters at time. For example, if both length and width increase or both decrease, fractal dimension may remain the same, as supported by the diagonal yellow trend.
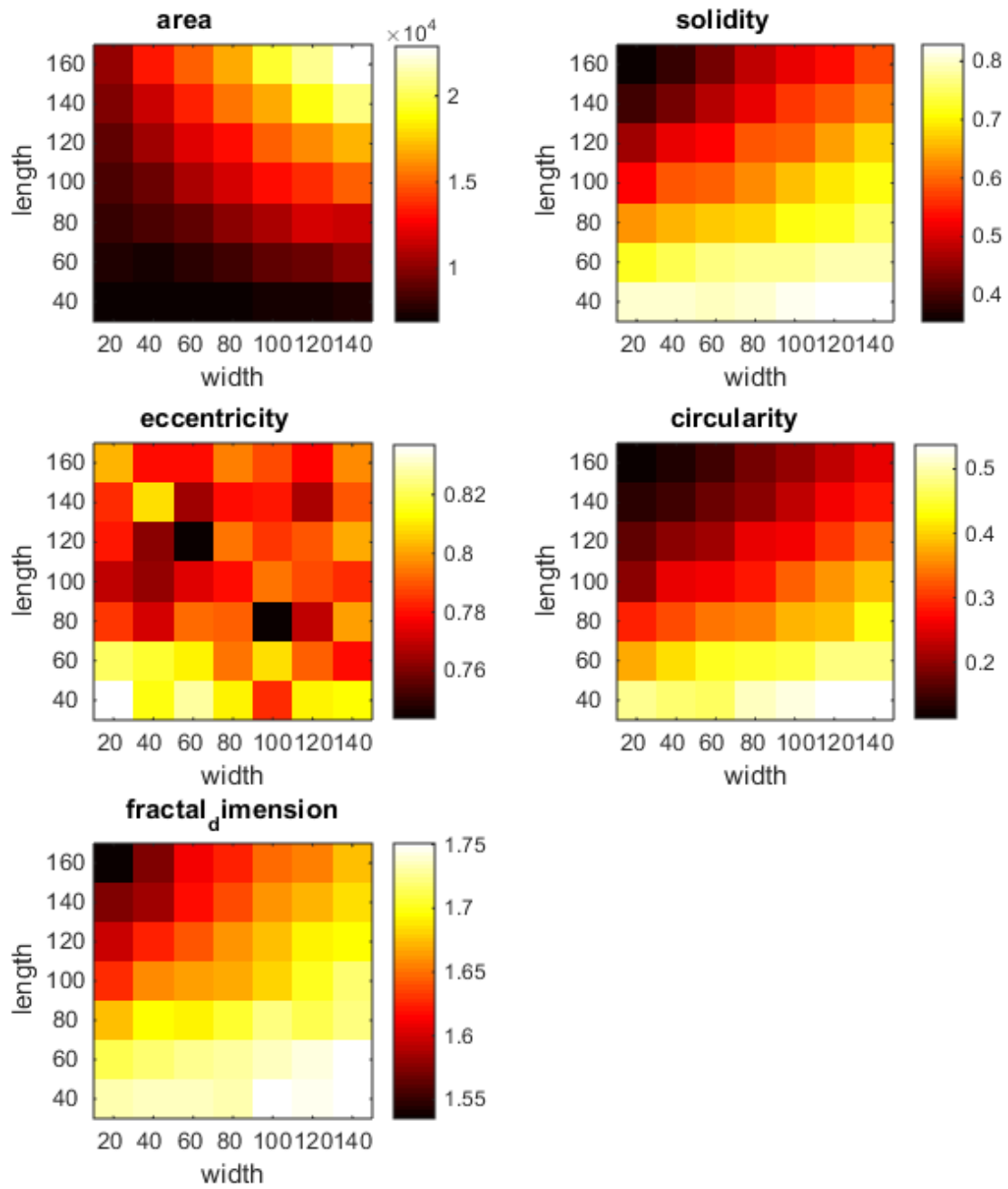
Figure 13: Two parameter heat Maps for cell length and width.
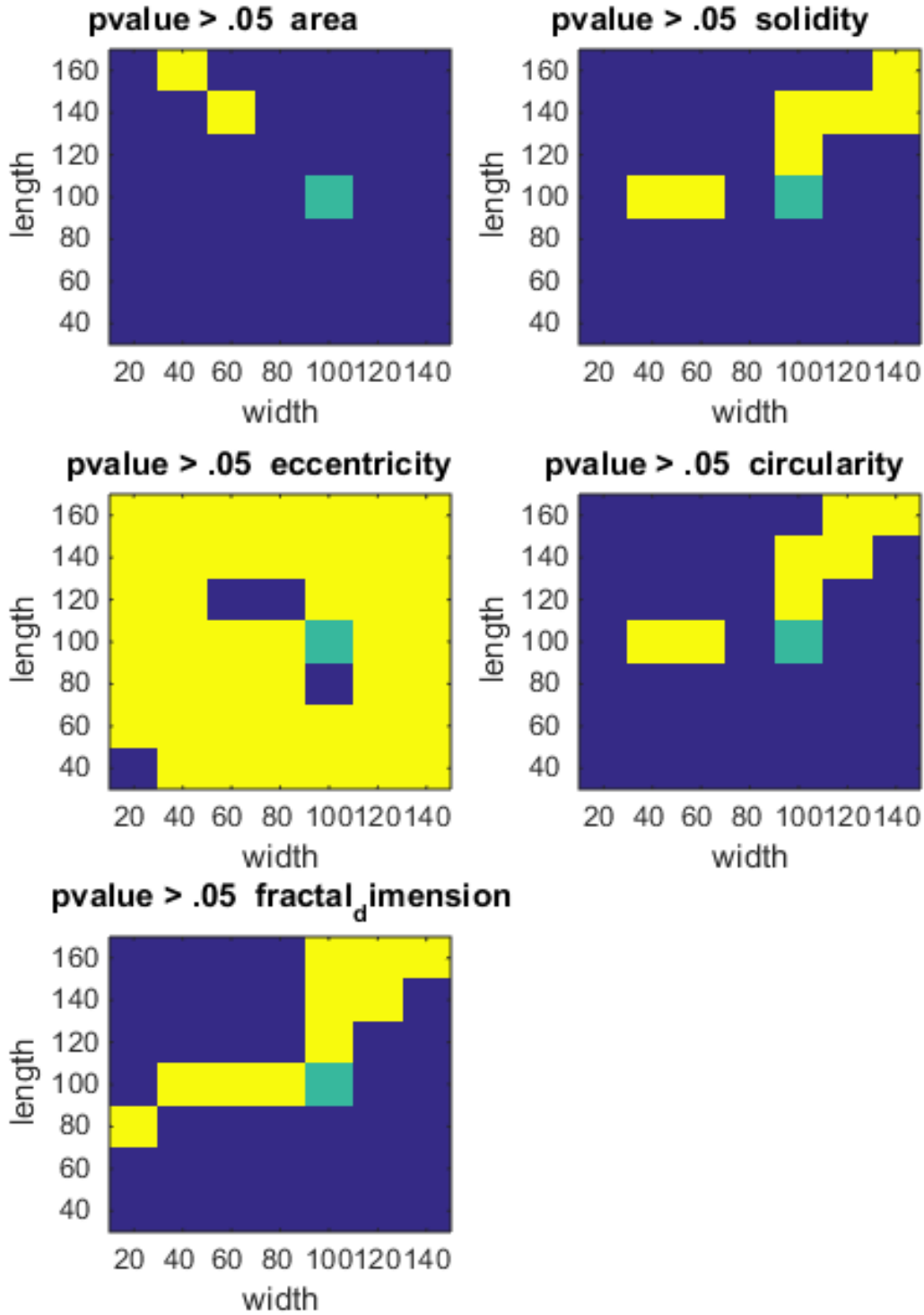
Figure 14: Two parameter T-test heat maps. Green represents the wild type and yellow means that the population is not statistically different from the wild type based on the metric.

Once I determined that some of our metrics did not provide information in differentiating populations with certain parameter variations, I wanted to determine which metrics provide information, and for which sets of parameters, in an automated way. In order to do this, I correlated every parameter vs. parameter heatmap for each metric with itself rotated by 180 degrees, as demonstrated in Figure 15. If there is a clear gradient in the heatmap, then the
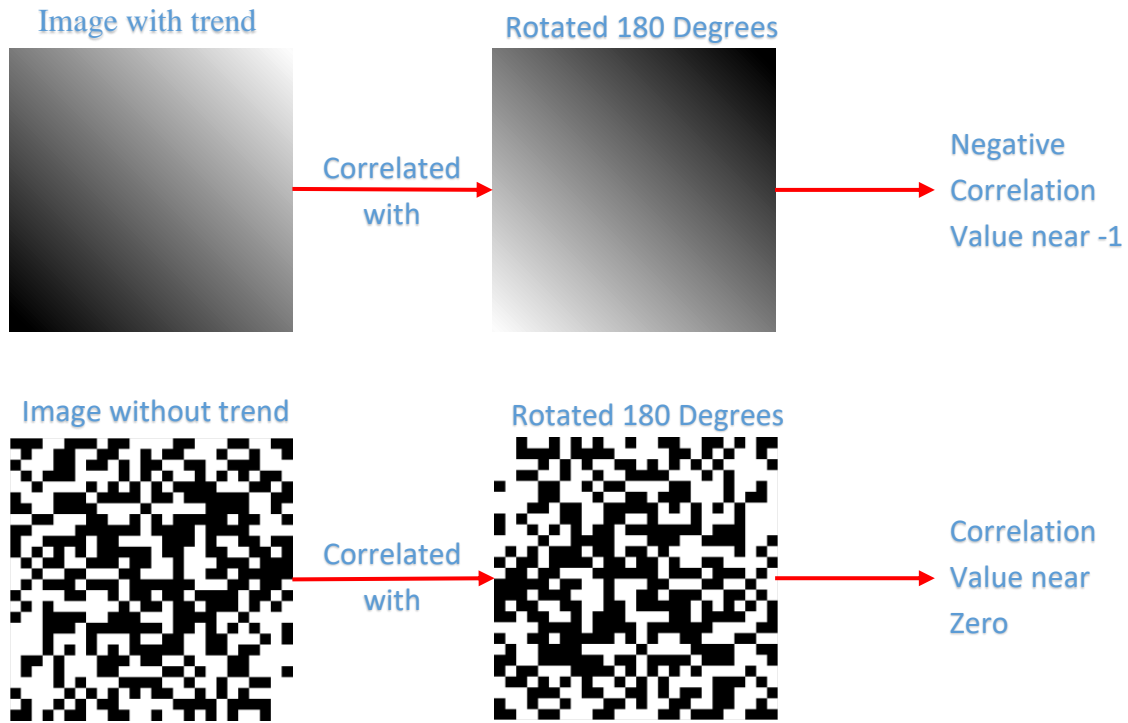
Figure 15: Example showing how Rotational Correlation is used to determine usefulness of a metric.

correlation operation should result in a value near negative one, since we are merely correlating with the gradient in the reverse direction. However, if the metric does not show a trend, such as the eccentricity heatmap shown in Figure 13, then, when we perform the rotational correlation, we will get a value near zero. I call this value for any given heatmap correlated with itself rotated by 180 degrees its rotational correlation (RC) value. The rotational correlation values for every

heatmap can be seen in Appendix 4. This test confirmed to us that mathematically eccentricity is

not useful at all in differentiating any of the parameters.
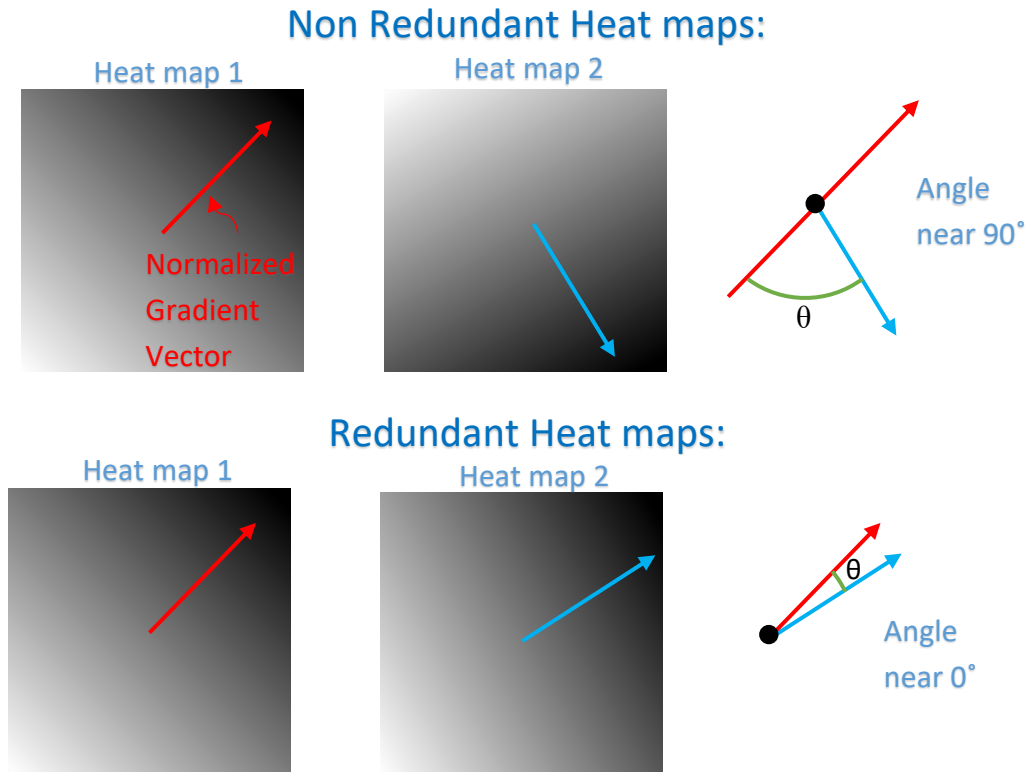
Figure 16: Example showing how the redundancy of heat maps can be assessed using angle between normalized direction vectors.

After I found out which metrics are useful for which pairs of parameters, I wanted to find

out which metric pairs are not redundant (show different trends). In order to do this, I calculated

the average direction of increase for the heat maps by finding the average gradient. Then, for

every metric pair, I calculated the angle between the average directional lines of the heat maps. If

the angle is near zero, then the metrics show the same trend and are redundant. However, if the

angle is near 90 degrees, then the metrics show different trends and together provide more

information. A demonstration of how this works is shown in Figure 16. This calculation allowed

me to remove all of the metrics pairs that did not provide addition information. Supported by

Figure 13, we see in Appendix 4 that for length and width, solidity, circularity, and fractal

dimension are all redundant (small angle) to eachother. Finally, I highlighted all the metric pairs

that showed useful and non-redundant trends in green in Appendix 4 and listed the useful and

non-redundant metrics pairs with respect to specific parameter pairs in Figure 17. These selected

metrics pairs could be useful in differentiating moss populations.

Parameter Pair	Useful and Non-Redundant Metric Pairs
length_branch	area_solidity
length_branch	area_fractal_dimension
length_branch	area_circularity
length_curve	area_circularity
curve_branch	area_circularity
length_curve	area_solidity
length_curve	area_fractal_dimension
length_curve	solidity_fractal_dimension
curve_branch	area_fractal_dimension
length_width	area_circularity
length_width	area_solidity
width_curve	area_fractal_dimension
length_width	area_fractal_dimension

Figure 17: Table listing statistically useful and non-redundant metric pairs with respect to specific parameter pairs

The early work on scatter plots, box graphs, and heat maps led to the rotational correlation

and the angle between heat maps analyses. These analyses showed that we have more

information on some parameters and very little on others. Looking at the table in Figure 17, we

see that we do not have many non-redundant metric pairs for any given pair of parameters. For

example, for the parameters length and branch we have three non-redundant metric pairs.

However, we learned that solidity, circularity, and fractal dimension are redundant metrics with

respect to length and branch, and therefore we actually only have one useful set for these

parameters. So, we need to find new metrics since we do not have sufficient information on certain parameter combinations.

3.4   Plant Structure Prediction

My ultimate goal is to be able to predict the structure of a plant just using metrics that can be obtained automatically. I do not believe that this can be achieved using a global model, one that has a single set formula, because moss growth is not completely deterministic. So, I needed to use a local model, one where the 'formula' changes based on the inputs. The model I chose was a K-nearest neighbors (KNN) model. K-nearest neighbors work by finding the k closest training set data points to the experimental data point. The classification of all the selected training set data points are averaged to give a prediction for the experimental point. My model postulates that the structure of an experimental plant will be similar to plants that have similar metric values. First, I trained the model (ExhaustiveSearcher in MATLAB) with means of the metrics of the 2,401 populations, discussed in Section 2.4, to get one data point per population. Then, when testing an experimental plant, I first calculated its metrics and found the k (a constant) closest data sets to that plant by its metrics (using knnsearch in MATLAB). I then averaged the structures of those data sets to get a prediction of the structure of the experimental plants (see code in Appendix 5). An example of this model with k=3 is shown in Figure 18.
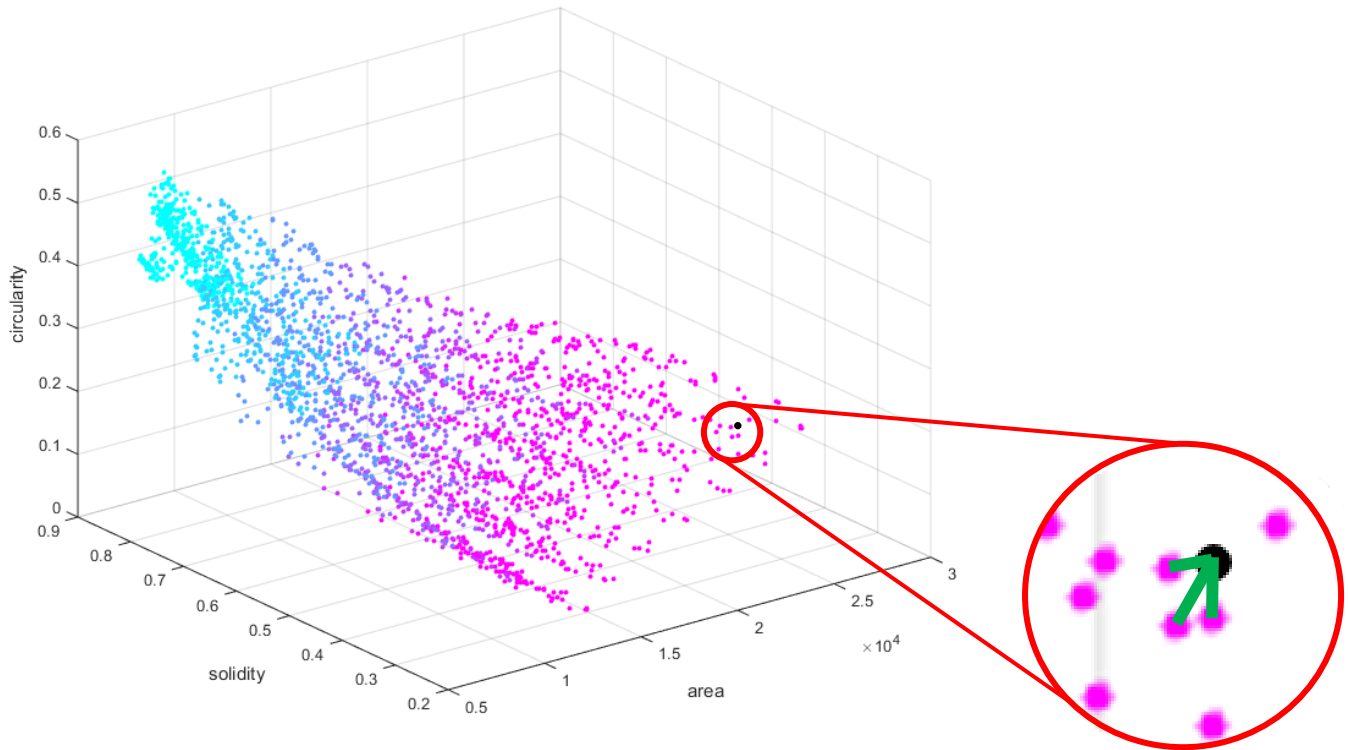
Figure 18: Visualization of k-nearest neighbors for k=3 in the data cloud.

In order to test how well the model works using our current metrics, I ran the model using all 100 plants of the wild type simulation population. If my model is accurate given the current set of metrics, I would expect that for each of the plants the output would be the wild type or at least that the mean of the output would be the wild type. When I used this testing procedure, I obtained an error of around 20% for each of the four parameters. I expected a large error rate because I have shown that the metric set is not currently very useful. More troubling is the fact that the error is essentially compounded four times, once for every parameter, making the prediction as a whole worse than 20%. Even after trying the model using different formulas for calculating the distances between neighbors, there was little improvement on the accuracy.

Although I did not end up testing it, the model may perform better if eccentricity is removed since, as I have shown in Section 3.3, that metric is useless.

4 Discussion

This project had two distinctly different sub-problems, an extended time imaging problem and a computational morphological assay problem. The goal of the first sub-problem was to create a method for imaging the growth of moss in three dimensions over time. I was successfully able to achieve this goal. The goal of the second sub-problem was to evaluate the effectiveness of our morphological assay and to explore how it could be used to predict experimentally determined plant morphology.

4.1 Time-Lapse Three-Dimensional Imaging of Moss Protonemata

In the majority of previous publications, only two dimensional imaging methods of moss protonemata have been used (Doonan et al., 1988; Vidali and Bezanilla, 2012; Bibeau et al., 2014; Ding et al., 2014). Having the ability to look at the third dimension allows questions to be answered that would not be easy to answer with two dimensional methods. For example, in two dimensional images, we cannot resolve the depth information in overlaps of plant branches. Furthermore, the capacity to image for an extended period of time provides temporal information, which is necessary to develop models that accurately represent growth.

The capacity to continuously image protonemata, will allow us to learn more about moss growth, including information such as how protonemata grow towards or away from light and how protonemal filaments interact between themselves. For example, we can now determine if there is a threshold of light intensity when the plants stop growing towards the light and start growing away from it, which may help determine the amount of light intensity for optimal growth. This may be complicated by the blue light used to image the plant, but because imaging

only takes place every 15 or 30 minutes, the imaging light dosage can be minimized to avoid any effect on growth. Nevertheless, work still needs to be done on how to handle the fact that calcofluor dye fades over time; for example, increased camera sensitivity and reduced exposure times should improve the dye bleaching problem.

It is important that we now have the capability to study additional questions, such as if the moss plant branches actively try to avoid each other or if they will grow into each other. This may provide valuable information on mechanisms used by the plant to achieve better light absorption. Future work can also be done with regard to mutation characterization. For example, mutations that interfere with plant growth can be studied continuously in three dimensions. This will provide information about how such mutations may change plant structure during growth, which could include changes in cell length, cell width, curving angle, branching angle, cell division patterns, and self-avoidance.

The methodology described in this work could be applied to other simple plants, and the knowledge derived may be valuable to understand more complex crop plants. This knowledge should help us find conditions as well as cellular and molecular mechanisms that are favorable to plant growth. This new methodology provides researchers with another tool to enhance their work in plant growth, thus, in the long term, aiding the movement toward agricultural sustainability.

4.2   Exploring the Parameter Space of a Computational Model of Protonemal Growth

Prior to this work, we did not know how useful our morphological assay was at quantifying moss plants. As such, I needed to develop methods and techniques to properly evaluate effectiveness of the metrics. In order to do this, I started by looking at scatter plots, box

graphs and heat maps to look for trends in the following metrics: area, solidity, circularity, eccentricity, and fractal dimension. I quickly learned that I would need a better method to quantify the usefulness of the metrics. I then developed the rotational correlation analysis using the same metrics to determine which metrics showed trends and which metrics did not show trends. The result of this rotational correlation analysis showed that eccentricity was not predictive. I then developed the angle between heat maps analysis and found that some of the other metrics were redundant, as supported by Appendix 4. Overall, our various analyses showed that the morphological assay does not capture information sufficient enough to differentiate changes in plant structure. As such, a better set of metrics needs to be found. I think a possible source of additional metrics can be found based on graph theory to help quantify the structure of a moss plant. For these metrics to work, each pixel of the plant could be considered to be a node. Possible graph theory metrics could include graph interconnectivity and node degree (Lund et al., 2009).

Although we know that with our current set of metrics we cannot accurately detect the level of differences when a single parameter is changed, we may be able to tell if two populations are different with 95% confidence using the T-test method that was described in Section 3.3. This would work best using the metric pairs which are highlighted in green in Appendix 4 because I have determined that these pairs show non-redundant trends. This method could then be used to experimentally screen if a mutant plant or a different moss strain is different morphologically, not how it is different, suggesting that deeper analysis needs to be performed.

In an attempt to predict experimental plant morphology, I developed a K-nearest neighbors model, as described in Section 3.4. My study showed that this model is not currently able to predict growth parameters based on our current metrics, since the metrics do not provide enough distinguishing information. Therefore, further work needs to be done to create a stronger predictive model. For a predictive model to be useful, a few improvements need to be made to bring the accuracy up to a reasonable level (see suggestions above on improving metrics). In general, having additional and better metrics can improve the performance of the prediction of the model. I anticipate that, once we have a good set of metrics, we can produce good estimates for the structure of experimental plants using this model. Also, we cannot be certain about the testing of the accuracy of the model. This is because it is possible that some of the simulated plants are mislabeled, as belonging to the wrong population, since the moss growth simulator allows some variation. Therefore, mislabeling of training data is likely to cause lower than expected accuracy for the model.

To improve our KNN model, it would help to sharpen the data points used to train the model. The plants in each population have variation and, to reduce the effect of this variation on the data point generated from it, we could increase the number of plants in the population or decrease the amount of allowed variation in each plant. Another issue I discovered that can affect the predictive power of the KNN approach is that moss of the same population can have a different number of starting bulbs. The moss grows from these bulbs and plants with more starting bulbs grow more despite being from the same population. I think that it would help to separate the populations into different data points based on the number of starting bulbs.

It is relevant to mention that both the morphological assay and the KNN model do not require any human input, which means that we can use it to quickly screen thousands of different plants with different mutations to find mutant plants and also learn more about these mutations. I believe that using our three dimensional extended time microscopy method in conjunction with our KNN prediction model, with the mentioned improvements to the model and metrics, we will be able to conduct accurate and fast screening of experimental moss plant populations. We then may be able to find genetic variations that lead to favorable plant growth.

References

Bibeau, J.P., Vidali, L., Zárský, V., and Cvrcková, F. (2014). Morphological Analysis of Cell Growth Mutants in *Physcomitrella*. In Plant cell morphogenesis : methods and protocols, pp. 201-213.

Deal, S., Wambaugh, J., Judson, R., Mosher, S., Radio, N., Houck, K., and Padilla, S. (2016). Development of a quantitative morphological assessment of toxicant-treated zebrafish larvae using brightfield imaging and high-content analysis. J Appl Toxicol **36,** 1214-1222.

Ding, X.S., Vidali, L.F.a.B.B., and Flaherty, P.J.F.a.B.E. (2014). Isolation of Temperature-sensitive Mutants in the Moss *Physcomitrella patens* and Mapping of their Causal Mutation by Genome Sequencing of Pooled Segregants. (Worcester Polytechnic Institute).

Domozych, D.S. (2012). The quest for four-dimensional imaging in plant cell biology: it's just a matter of time. Ann Bot **110,** 461-474.

Doonan, J.H., Cove, D.J., and Lloyd, C.W. (1988). Microtubules and microfilaments in tip growth: evidence that microtubules impose polarity on protonemal growth in *Physcomitrella patens*. Journal of Cell Science **89,** 533.

Eckardt, N.A. (2008). High-resolution three-dimensional imaging of plant tissues. The Plant Cell **20,** 1423-1423.

FAO. (2013). FAO statistical yearbook 2013 world food and agriculture (FAO), pp. xvi-285.

Flavell, R. (2009). Role of model plant species. Methods Mol Biol **513,** 1-18.

Knight, C.D., and Perroud, P.-F. (2010). *Physcomitrella patens*: A Model Bryophyte, eLS

Lemoi, K.P.S.a.P.H., and Tüzel, E.F.a.P.H. (2012). Biophysical Insights into Moss Morphology and Organelle Motility. (Worcester Polytechnic Institute).

Lund, A.W., Bilgin, C.C., Hasan, M.A., McKeen, L.M., Stegemann, J.P., Yener, B., Zaki, M.J., and Plopper, G.E. (2009). Quantification of Spatial Parameters in 3D Cellular Constructs Using Graph Theory. Journal of Biomedicine and Biotechnology **2009,** 16.

McNally, J.G., Karpova, T., Cooper, J., and Conchello, J.A. (1999). Three-Dimensional Imaging by Deconvolution Microscopy. Methods **19,** 373-385.

Murphy, D.B., and Davidson, M.W. (2012). Fundamentals of Light Microscopy and Electronic Imaging. (Wiley-Blackwell).

Prigge, M.J., and Bezanilla, M. (2010). Evolutionary crossroads in developmental biology: *Physcomitrella patens*. Development **137,** 3535-3543.

Rounds, C.M., Lubeck, E., Hepler, P.K., and Winship, L.J. (2011). Propidium Iodide Competes with Ca(2+) to Label Pectin in Pollen Tubes and Arabidopsis Root Hairs. Plant Physiology **157,** 175-187.

Schaefer, D.G., and Zrÿd, J.-P. (2001). The Moss *Physcomitrella patens*, Now and Then. Plant Physiology **127,** 1430-1438.

Schmid, B., Schindelin, J., Cardona, A., Longair, M., and Heisenberg, M. (2010). A high-level 3D visualization API for Java and ImageJ. BMC Bioinformatics **11,** 274-274.

Stefan, B., and Claude, L. (2007). Evolution of Virtual Plants Interacting with Their Environment (ResearchGate).

Steinberg, D., and Sikora, S. (1999). Plant growth simulation in virtual worlds : towards online artificial ecosystems. In Conference on Artificial Life, pp. 19-25.

ThermoFisher. (2009). Fluorescent probes for plant imaging (Invitrogen).

Truernit, E., Bauby, H., Dubreucq, B., Grandjean, O., Runions, J., Barthelemy, J., and Palauqui, J.C. (2008). High-resolution whole-mount imaging of three-dimensional tissue organization and gene expression enables the study of Phloem development and structure in Arabidopsis. Plant Cell **20,** 1494-1503.

Vidali, L., and Bezanilla, M. (2012). *Physcomitrella patens*: a model for tip cell growth and differentiation. Curr Opin Plant Biol **15,** 625-631.

Appendix

Appendix 1: Batch file to generate the simulated plant populations.

```
01 setlocal enableDelayedExpansion
02
03 FOR %%A IN (20 40 60 80 100 120 140) DO (
04    for /f %%b in ('cscript //nologo eval.vbs %%A 0.71835') do (set
"len=%%b") rem Multiply Length percentage by experimental
05    for /f %%b in ('cscript //nologo eval.vbs !len! 0.308') do (set
"lenstd=%%b") rem Keep the length standard deviation as 31% of length
06    FOR %%B IN (40 60 80 100 120 140 160) DO (
07        for /f %%b in ('cscript //nologo eval.vbs %%B 0.1006') do (set
"wid=%%b") rem Multiply Length percentage by experimental
08        FOR %%C IN (1 15 30 45 60 75 90) DO (
09            for /f %%b in ('cscript //nologo eval2.vbs %%C') do (set
"cur=%%b") rem Multiply Curving percentage by experimental
10            FOR %%D IN (0 15 30 45 60 75 90) DO (
11                for /f %%b in ('cscript //nologo eval.vbs %%D 1.0') do (set
"bra=%%b") rem Multiply Branching percentage by experimental
12                rem Run the simulation with the new parameters
13                java -jar mossgrowth2.jar 1200 1200 4 100 !len! !lenstd!
!wid! 0.0 !cur! !bra! 16.617 1.2 0 0 24.546 0.0 0.9
length%%Awidth%%Bcurve%%Cbranch%%D
14            )
15        )
16    )
17 )
18 pause
```
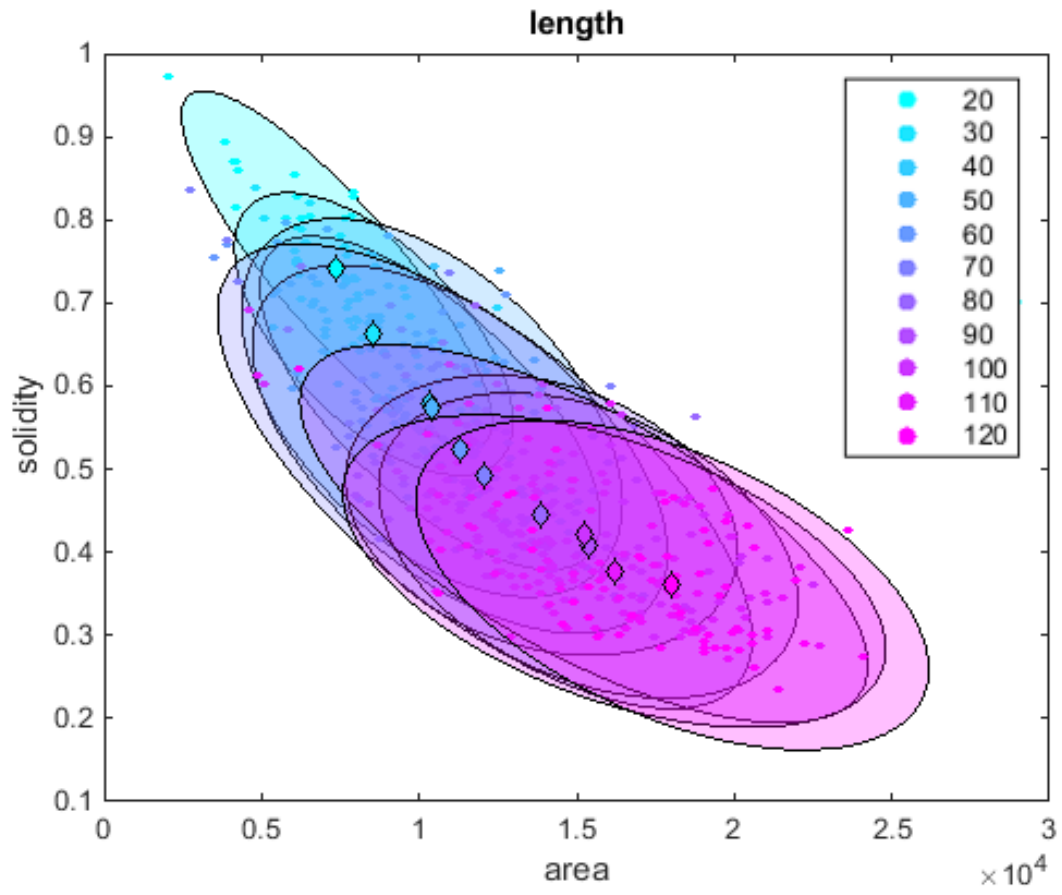
Appendix 3: Code for plotting two metrics against each other with one parameter and plotting with PCA.

```matlab
01 %Generate figure showing solidity vs area colored by length
02 figure
03 para='length'
04 for j=20:10:121
05     data=load(strcat(para,int2str(j)));
06     col=[(j-20)/100,1-(j-20)/100,1];
07     scatter(data.results(4,:,1),(data.results(4,:,2)),10,col,'filled')
08     hold on
09 end
10 xlabel('area');
11 ylabel('solidity');
12 legend(strread(num2str(20:10:121),'%s'),'Location','northeast')
13 title(para)
14 %Plot means of each population
15 for j=20:10:121
16     data=load(strcat(para,int2str(j)));
17     col=[(j-20)/100,1-(j-20)/100,1];
18
s=scatter(mean(data.results(4,:,1)),mean((data.results(4,:,2))),40,col,'d','filled');
19     s.MarkerEdgeColor = [0 0 0];
20     hold on
21 end
22
23 %Generate figure showing first two principal components colored by length
24 figure
25 para='length'
26 points=[];
27 %Get all data points from all populations
28 for j=20:10:121
29     data=load(strcat(para,int2str(j)));
30     points=cat(2,points,data.results(4,:,:));
31 end
32 points=squeeze(points);
33 %Perform pca on all points
34 [coeff,score,latent,tsquared,explained,mu] = pca(points);
35 count=0;
36 for j=20:10:121
37     col=[(j-20)/100,1-(j-20)/100,1];
38
scatter(score(count*50+1:count*50+50,1),score(count*50+1:count*50+50,2),10,col,'filled')
39     hold on
40     count=count+1;
41 end
42 xlabel('1st PC');
43 ylabel('2nd PC');
44 legend(strread(num2str(20:10:121 + '%'),'%s'),'Location','southwest')
45 title(para)
46 %Plot means of each population
47 count=0;
48 for j=20:10:121
49     col=[(j-20)/100,1-(j-20)/100,1];
50
s=scatter(mean(score(count*50+1:count*50+50,1)),mean((score(count*50+1:count*50+50,2))),40,col,'d','filled');
51     s.MarkerEdgeColor = [0 0 0];
52     hold on
53     count=count+1;
54 end
```

Appendix 4: Spreadsheet showing quantification of metric pairs.

Metric Pair	Rotational Correlation(RC) 1	RC 2nd Metric	Metric 1 has trend?	Metric 2 has trend?	Both?	Angle
length_branch_area_solidity	-0.982532513	-0.985521384	TRUE	TRUE	TRUE	53
length_branch_solidity_circularity	-0.985521384	-0.935179186	TRUE	TRUE	TRUE	1
width_branch_solidity_circularity	-0.977028688	-0.977375213	TRUE	TRUE	TRUE	5
length_width_solidity_circularity	-0.927653055	-0.943865449	TRUE	TRUE	TRUE	2
length_branch_solidity_fractal_dimension	-0.985521384	-0.965711579	TRUE	TRUE	TRUE	3
length_width_solidity_fractal_dimension	-0.927653055	-0.85300699	TRUE	TRUE	TRUE	9
length_branch_area_fractal_dimension	-0.982532513	-0.965711579	TRUE	TRUE	TRUE	53
length_branch_area_circularity	-0.982532513	-0.935179186	TRUE	TRUE	TRUE	53
length_branch_circularity_fractal_dimension	-0.935179186	-0.965711579	TRUE	TRUE	TRUE	5
width_branch_area_fractal_dimension	-0.969367558	-0.939056908	TRUE	TRUE	TRUE	9
width_branch_solidity_fractal_dimension	-0.977028688	-0.939056908	TRUE	TRUE	TRUE	9
length_width_circularity_fractal_dimension	-0.943865449	-0.85300699	TRUE	TRUE	TRUE	11
curve_branch_eccentricity_fractal_dimension	-0.372984906	-0.705188126	FALSE	TRUE	FALSE	53
length_curve_area_circularity	-0.981682957	-0.91313257	TRUE	TRUE	TRUE	53
width_branch_circularity_fractal_dimension	-0.977375213	-0.939056908	TRUE	TRUE	TRUE	14
width_branch_area_solidity	-0.969367558	-0.977028688	TRUE	TRUE	TRUE	18
curve_branch_area_circularity	-0.840081394	-0.953296661	TRUE	TRUE	TRUE	53
length_curve_area_solidity	-0.981682957	-0.760225037	TRUE	TRUE	TRUE	53
width_curve_circularity_fractal_dimension	-0.951192168	-0.764945175	TRUE	TRUE	TRUE	29
length_curve_solidity_circularity	-0.760225037	-0.91313257	TRUE	TRUE	TRUE	37
width_branch_area_circularity	-0.969367558	-0.977375213	TRUE	TRUE	TRUE	24
length_curve_circularity_fractal_dimension	-0.91313257	-0.851243365	TRUE	TRUE	TRUE	30
curve_branch_circularity_fractal_dimension	-0.953296661	-0.705188126	TRUE	TRUE	TRUE	27
width_curve_eccentricity_fractal_dimension	-0.395897954	-0.764945175	FALSE	TRUE	FALSE	53
length_curve_area_fractal_dimension	-0.981682957	-0.851243365	TRUE	TRUE	TRUE	53
width_curve_solidity_circularity	-0.529155535	-0.951192168	FALSE	TRUE	FALSE	60
width_curve_area_circularity	-0.977199135	-0.951192168	TRUE	TRUE	TRUE	30
length_curve_solidity_fractal_dimension	-0.760225037	-0.851243365	TRUE	TRUE	TRUE	67
curve_branch_eccentricity_circularity	-0.372984906	-0.953296661	FALSE	TRUE	FALSE	53
curve_branch_area_solidity	-0.840081394	-0.269046132	TRUE	FALSE	FALSE	68
length_curve_eccentricity_fractal_dimension	-0.54974592	-0.851243365	FALSE	TRUE	FALSE	55
length_branch_eccentricity_circularity	-0.437818145	-0.935179186	FALSE	TRUE	FALSE	13
curve_branch_area_fractal_dimension	-0.840081394	-0.705188126	TRUE	TRUE	TRUE	53
width_curve_area_solidity	-0.977199135	-0.529155535	TRUE	FALSE	FALSE	30
curve_branch_solidity_circularity	-0.269046132	-0.953296661	FALSE	TRUE	FALSE	53
length_branch_solidity_eccentricity	-0.985521384	-0.437818145	TRUE	FALSE	FALSE	14
length_branch_area_eccentricity	-0.982532513	-0.437818145	TRUE	FALSE	FALSE	53
length_width_area_circularity	-0.791628431	-0.943865449	TRUE	TRUE	TRUE	56
length_width_eccentricity_circularity	0.17789324	-0.943865449	FALSE	TRUE	FALSE	26
length_width_area_solidity	-0.791628431	-0.927653055	TRUE	TRUE	TRUE	58
length_width_solidity_eccentricity	-0.927653055	0.17789324	TRUE	FALSE	FALSE	28
width_curve_solidity_fractal_dimension	-0.529155535	-0.764945175	FALSE	TRUE	FALSE	90
length_branch_eccentricity_fractal_dimension	-0.437818145	-0.965711579	FALSE	TRUE	FALSE	18
width_curve_eccentricity_circularity	-0.395897954	-0.951192168	FALSE	TRUE	FALSE	68
width_curve_area_fractal_dimension	-0.977199135	-0.764945175	TRUE	TRUE	TRUE	59
length_width_eccentricity_fractal_dimension	0.17789324	-0.85300699	FALSE	TRUE	FALSE	38
curve_branch_area_eccentricity	-0.840081394	-0.372984906	TRUE	FALSE	FALSE	51
length_width_area_fractal_dimension	-0.791628431	-0.85300699	TRUE	TRUE	TRUE	67
length_width_area_eccentricity	-0.791628431	0.17789324	TRUE	FALSE	FALSE	53
width_branch_eccentricity_circularity	-0.506280101	-0.977375213	FALSE	TRUE	FALSE	51
width_branch_solidity_eccentricity	-0.977028688	-0.506280101	TRUE	FALSE	FALSE	56
length_curve_solidity_eccentricity	-0.760225037	-0.54974592	TRUE	FALSE	FALSE	58
width_curve_area_eccentricity	-0.977199135	-0.395897954	TRUE	FALSE	FALSE	82
curve_branch_solidity_fractal_dimension	-0.269046132	-0.705188126	FALSE	TRUE	FALSE	53
length_curve_eccentricity_circularity	-0.54974592	-0.91313257	FALSE	TRUE	FALSE	85
width_curve_solidity_eccentricity	-0.529155535	-0.395897954	FALSE	FALSE	FALSE	52
curve_branch_solidity_eccentricity	-0.269046132	-0.372984906	FALSE	FALSE	FALSE	17
length_curve_area_eccentricity	-0.981682957	-0.54974592	TRUE	FALSE	FALSE	85
width_branch_area_eccentricity	-0.969367558	-0.506280101	TRUE	FALSE	FALSE	74
width_branch_eccentricity_fractal_dimension	-0.506280101	-0.939056908	FALSE	TRUE	FALSE	65

Both metrics have trends (threshold ABS(RC)>0.6)
Both metrics have trends and are not redundant (based on angle)

Appendix 5: Code for KNN model.

```matlab
01 %Builds the KNN model
02 k=20; %Number of neighbors to consider
03 testpoint=200000; %Point to check accuracy with
04 data=load(strcat('CondensedData\data.mat'));
05 data=data.data;
06 datatofind = data(testpoint,:)
07 target=load(strcat('CondensedData\target.mat'));
08 target=target.target;
09 expectedtarget=target(testpoint,:)
10 model = ExhaustiveSearcher(data,'Distance','cosine');%Build efficient
search tree (distance heuristic can be changed)
11 display('Model Built')
12 ind = knnsearch(model,datatofind,'K',k);%Find indices of k nearest
neighbors
13 knn=target(ind(:),:);%Get values of the neighbors
14 pred = mean(knn);%Prediction is the mean of the neighbors
15
display(strcat('length',int2str(pred(1)),'width',int2str(pred(2)),'curve',int
2str(pred(3)),'branch',int2str(pred(4))))
16 RMSE = sqrt(mean((expectedtarget-pred).^2))%Root Mean Squared Error
17 diff =expectedtarget-pred %Difference
18 perdiff = (abs(expectedtarget-pred)./expectedtarget)*100 %Percent
Distance1
```