University of Nevada, Reno

A simple and reliable method for estimating cactus pear biomass production using cladode dimensions

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in
Biochemistry

by

Gabriel J. Angres

Dr. John C. Cushman/Thesis Advisor

December 2022

UNIVERSITY OF NEVADA, RENO

THE GRADUATE SCHOOL

We recommend that the thesis prepared under our supervision by

GABRIEL J ANGRES

entitled

"A simple and reliable method for estimating cactus pear biomass production using cladode dimensions"

be accepted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

Dr. John C. Cushman
Advisor

Dr. Won C. Yim
Committee Member

Dr. Juan Soloman
Graduate School Representative

Markus Kemmelmeier, PhD., Dean
Graduate School

December 2022

Abstract

In a warming world, crassulacean acid metabolism (CAM) plants, which exhibit much greater water-use efficiency and tolerance to high temperatures than C_3 and C_4 photosynthesis plants, are increasingly attractive as food, feed, fiber, and biofuel crops. Hence, as agricultural development of these plants expands, the issue of measuring crop yield becomes more relevant. Traditionally, biomass estimates of a crop are based upon gravimetric assessment; however, this process is expensive, labor-intensive, and time-consuming. Indirect approaches that make accurate predictions from dimensional measurements, images, or simple object counts are simpler, faster, and relatively inexpensive. However, current indirect methods for photosynthetic-stem (cladode) plants, such as cactus pear (*Opuntia* spp.) require further refinement to be used widely in field settings. An optimal model for indirect measurement must be parsimonious and highly accurate, while also having a cogent supporting rationale for the model's construction. We conducted statistical analyses on cladode dimensional data using 14 accessions representing five species, including *Opuntia cochenillifera*, *O. crassa*, *O. ficus-indica*, *O. robusta*, and *O. undulata* to derive a general model for approximating the fresh weight of a cladode from its physical dimensions, and quantify any statistically significant morphological differences among them. Correlations between fresh weight and four physical measures (pad *diameter*, *width*, *height*, and *thickness*) and factorial combinations thereof (e.g., *height* x *width* vs. *fresh weight*) were assessed for accuracy (per the coefficient of determination) and parsimony (per the Schwarz-Bayes Criterion) (> 0.86 adj. R^2 minimum, 0.95 adj. R^2 general fit). A second modeling approach used a refined measured area in ImageJ (> 0.85 adj. R^2 minimum, 0.94 adj. R^2 general fit). A third modeling approach attempted an approximate but less logistically intensive area approximation using an ellipse which lost relatively little performance over all accessions (> 0.79 adj. R^2, 0.94 adj. R^2 general fit). These formulae generally met or exceeded the performance of models obtained from prior research efforts on an all-accessions basis. Further theoretical work consists of incorporating the effect of morphological variation and water content into the formula for a particular accession.

Keywords: cactus pear, productivity modeling, drought stress, linear regression

Acknowledgements

1 | Introduction

Over 41% of the global land area is arid (Adeel et al., 2005), and this area is predicted to increase by up to 23% within the next century due to global climate change (Huang et al., 2020). With this aridification, average temperatures will rise, diminishing the photosynthetic efficiency of major crops (Kirschbaum and McMillan, 2018). In addition, water availability will decrease (Şen, 2009), putting a further strain on agricultural productivity. The combined effect of these changes is that existing growing zones for established, widely cultivated crops will shift or disappear (Hijmans, 2003), significantly diminishing crop productivity. Therefore, efforts to adopt, develop, and establish "climate-ready" crops will receive substantial attention in the short and long term. Large agricultural CAM species, which favor growth in arid conditions, show great potential for such efforts. By fixing CO_2 at night instead of during the daytime, CAM plants reduce transpiration; as a result, those plants have a 6-fold and 4-fold greater water-use efficiency, respectively, compared to typical C_3 and C_4 photosynthetic crops currently in use (Borland et al., 2009). Many CAM species, including cactus pear (*Opuntia* spp.) are used extensively for food, forage, biofuel, and other secondary products (Davis et al., 2019).

Field-yield prediction is one of the central components of large-scale agriculture for both farmers and researchers. Farming operations benefit from crop yield prediction data to determine production quantity, harvest logistics, and profit (Agriculture Victoria, 2022) and to coordinate with governments (Organisation for Economic Co-operation and Development, 2022) and markets (Krejcia and Beamon, 2015). Meanwhile, agronomic researchers farming on smaller scales require field-yield predictions for projects that optimize agricultural output or analyze the effect of a given growing environment on production. Thus, the commercial and experimental development of a particular crop relies in part on accurate and reliable crop yield estimations. The two broad classes of field-yield evaluations are direct, destructive measurement and indirect, non-destructive measurement and modeling.

Destructive measurement uses conventional sampling methods and direct analysis of the sample's properties (Neupane et al., 2021). Destructive measurement approaches have high accuracy but can be

logistically costly for farmers because the techniques require labor-intensive sample collection, storage, transport, and disposal. For agronomic researchers, additional costs arise because destructive sampling alters a target plant's condition in the experiment and prevents any future sampling of the plant. To avoid using the same plants for multiple sampling periods, researchers either must limit their studies to a single growing season, or expand the growing plots and incur significant land, water, and soil fertilizer costs.

In contrast, indirect, non-destructive measurement and modeling approaches use measurable characteristics of the target plant to approximate the yield without removing or modifying any plant parts. Consequently, indirect measurement is less accurate and more approximate than destructive measurement but comes with significant logistical advantages. Indirect measurements have negligible effects on plant growth and can be used to obtain multiple sample observations from a single plant. Some examples of in-place measurements include measuring crop leaf area (Colaizzi et al., 2017) or light interception (Pinheiro et al., 2015). However, indirect models used in place of a direct measurement must be proven to be accurate and reliable.

In the past twelve years, four experimental groups have made progress towards establishing accurate and reliable indirect, non-destructive measurement and modeling schemes for *Opuntia* growing efforts. Collectively, their reports have established that indirect modeling in the cactus pear is feasible, but that there are several outstanding challenges to consistent model performance across multiple accessions.

Tiznado-Hernández et. al., 2010, sought to estimate *Opuntia* pad area using an ellipse with the same height and width as the target cladode. The authors produced highly accurate models for each of two accessions under testing "Copena F-1" and "Atlixco" with $R^2 > 0.94$ and $n = 80$ cladodes for each. Thus, the authors established that an ellipse is a reasonable approximation of pad area. However, a larger bias in estimation occurred between accessions using the elliptical approximation (Tiznado-Hernández et al., 2010).

Pinheiro et. al., 2015 tested several indirect measurement methods to accurately model the area index of fresh cactus cladodes. After testing various other techniques, including light interception methods, measurements through leaf disc samples, and planimetry, the authors found that pad area

estimation from one-dimensional (length) measurements produced reasonably accurate approximations (n = 180 plants). Using equations previously derived for the *Opuntia* accessions "IPA Sertânia", "Miúda" and "Orelha de Elefante Mexicana", the authors were able to predict area with 94%-99% accuracy using only the height, width, and perimeter of the cactus pad. Thus, the authors showed that accurate prediction of pad weight using pad area was possible. It should be noted, however, that "[model] applicability is dependent upon the use of estimated parameters or adjusted value equations", which means that tuned modeling equations developed for one accession may not necessarily apply to other varieties (Pinheiro et al., 2015).

Reis et. al., 2018 attempted to characterize the pad area, fresh weight, and dry weight of several *O. ficus-indica* accessions using relations to the width, length, thickness, and neck diameter of the cladodes. After testing thirty different models with an n = 60 cladodes, the authors established cladode area as a function of length and width with $R^2 = 0.91$, and fresh weight as a function of length, width, and thickness with an R^2 of 0.91. However, the authors noted that "the lack of variance homogeneity in four of the studied variables…reflects the existence of intra-population variability" – essentially establishing that such parameter models might not necessarily extend to other sample groups from the same species or variety. More importantly, although multiple models were tested, no mathematical rationale for any model was established, so no clear way to establish whether the fitted models have any generalizability between samples was evident (Reis et al., 2018).

Lucena et. al., 2021 attempted to model area and fresh weight of *Nopalea cochenillifera* using height, width, and thickness employed in three formula types (linear, power law, and gamma) for a total of six formula attempts (Lucena et al., 2021). Using a n = 432 cladodes, the authors determined that a power law for area (height and width raised to a constant power) and a gamma law (height times width plus thickness) for fresh weight had the highest performance at 99.97 R^2 minimum and 98.51 R^2 minimum, respectively, over all cladodes tested (Lucena et al., 2021). This effort appears to be the most successful single-accession indirect measurement project to date. However, once again, results derived for

a single accession cannot be generalized across all possible accessions.

Although some relationships between fresh weight, area, and length measurements in *Opuntia* have been established, an overall lack of clarity in modeling efforts remains. Although some studies acknowledge differences in model performance between accessions, no prior work has attempted to discern the causes of the differences among accessions. Conversely, no experiment has directly investigated the similarity among accessions. Logically, these differences and similarities must exist because no two accession-specific pad morphologies are wholly unique or wholly common to each other. Visually and intuitively, such differences are readily apparent in the field. Therefore, to achieve a thorough understanding of indirect modeling in *Opuntia*, a comprehensive study of modeling effectiveness must quantify the differences in shape variation among accessions *and* relate these differences to the variations in model performance among accessions. In addition, all models tested must be informed by mathematical reasoning represented by an abstract equation with a concrete interpretation that explains thoroughly the inclusion, arrangement of, and operations on each term.

2 | Methodology/Statistical Analysis

2.1 | Field Planting

Prior to field planting, pads from nine different accessions of *Opuntia ficus-indica*, three accessions of *O. cochinellifera*, one accession of *O. undulata*, and one accession of *O. robusta* from the UNR greenhouse located at the Valley road field lab, Reno, NV, and from a collection of adult plants from the USDA National Arid Land Germplasm Resource Unit (NALPGRU), Parlier, CA were removed from adult plants and given two weeks to callus under shade to promote callus formation and prevent microbe infection from immediate contact with soil in the field (Neupane et al., 2023). Pads were planted in the USDA NALPGRU at N 36° 35' 36.5" (latitude), W 119° 30' 52.6" (longtitude), in soil categorized as Hanford fine sandy loam (USDA Natural Resources Conservation Service, 2022). Plants were planted on March

18th, 2019, in a pseudo-randomized block design with eight plants per block, with four replicated blocks per accession, for a total of 32 plants per accession. The field layout used a spacing of three m between rows and 0.5 m between plants within rows. Additional *O. robusta* was planted at either end of each row and in separate rows that flanked the first and last rows to buffer any possible edge effects. Plants were irrigated using pressure-compensating drip emitters on irrigation lines that provided a calculated 800 mm y^{-1} in addition to the average annual irrigation of 312 mm y^{-1}. Fertilizer was applied uniformly through the drip irrigation system on a monthly basis using All Purpose LiquaFeed® Miracle Gro® fertilizer N:P:K 12:4:8 (Scott's Miracle Gro, Inc., Marysville, OH, USA) according to manufacturer's instructions. Representative images of the 14 *Opuntia* accessions used are presented in Figure 1.

Figure 1. Representative images of 14 elite *Opuntia accessions* used for the current field trial located at National Arid Land Plant Genetic Resource Unit (NALPGRU) at Parlier, CA, from 2019-2022. Accessions included *O. ficus-indica* (PARL 246, PARL 572, PARL 845, PARL 854, PARL 584, PARL 580, and PARL 585), *O. crassa* (PARL 325), *O. cochenillifera* (PARL 242, PARL 319, PARL 582), O. robusta (PARL 390), and *O. undulata* (PARL 839). Size bar = 10 cm.

2.2 | Measurement Preparation

After field planting and cultivation, selected terminal pads were taken from individual trees, and measurements of cladode *height*, *width*, *thickness*, *fresh weight*, and *diameter* were taken. For each measure:

- *Height* is the length of the pad,

- *Width* is the longest length of the pad as taken on an orthogonal axis in the same plane as the *height*,

- *Thickness* is the length of the cactus measured on the axis orthogonal to the plane of the axes of *height* and *width*,

- *Fresh Weight* is the weight of the pad as recorded on a scale,

- *Diameter* is the *width* of the circle that has the same perimeter as the pad.

The *diameter* measure may be considered an analogue of the measure of "perimeter" in Pinheiro et. al. (2015). The *diameter* measure has several mathematical properties relevant to the determination of *fresh weight*. For instance, it can be trivially established that, for any cladode,

EQ 1

$$Height \geq Diameter \geq Width$$

Further, visual observation indicates that *diameter* varies proportionally with both *height* and *width* (Figure 2). Hence, diameter was included as part of the measure analysis going forward.
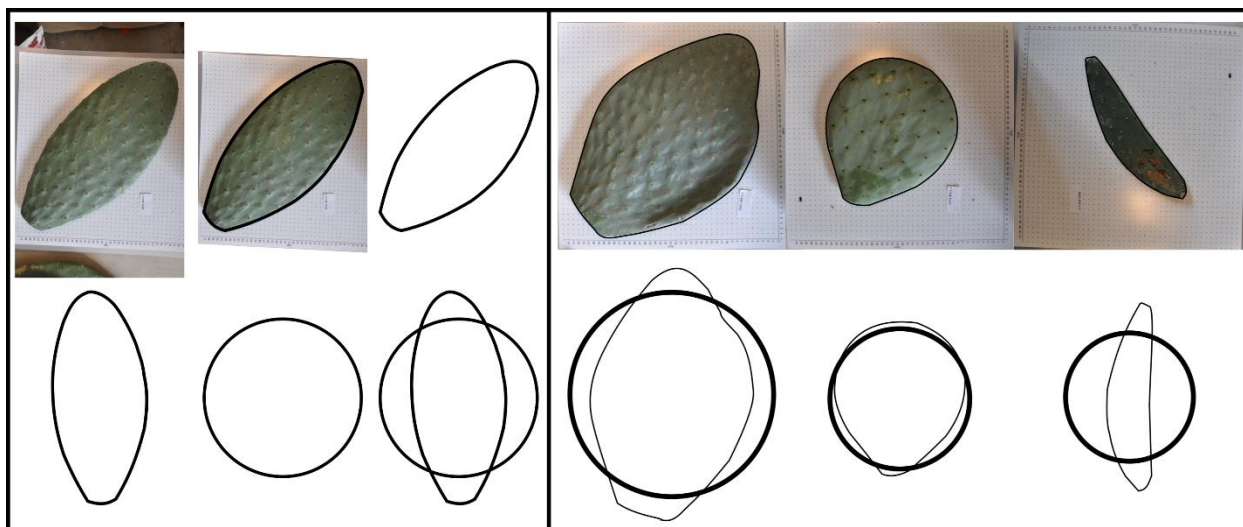
Figure 2. a) A depiction of the geometric process used to obtain the diameter measure. After obtaining the perimeter of the cactus (black outline), a circle with the equivalent perimeter is constructed, and the diameter of the resulting circle is included as a measure. b) A depiction of how the diameter measure encodes information about cactus pad shape in relation to pad width and height. As pad shape varies, the difference between diameter vs. height and diameter vs. width changes accordingly.

2.3 | Determining Testable Models

The basic equation for the mass of an object can be stated as:

EQ 2

$$Fresh\ Weight\ =\ f(Volume, Density)$$

Where *f(Volume, Density)* is the operation that determines the relationship of Volume and Density to the *fresh weight* from the dimensions of the shape under analysis,.

The density throughout the pad can be held constant, so any approximation of volume is by extension an approximation of *fresh weight*.

EQ 3

$$Fresh\ Weight\ =\ f(Volume) * density = Volume(measures) * density$$

From here, the formula will be presented as a proportional relation (with ≈), rather than an exact equivalence relation (=), to focus on the relation between the quantities.

EQ 4

$$Fresh\ Weight\ \approx Volume(measure_1, measure_2, measure_3, \cdots measure_n)$$

An initial starting point for estimating the volume of a cladode is using the volume of the smallest box

(rectangular prism) that the cladode may fit in – written as:

EQ 5

$$Fresh\ Weight \approx Height \cdot Width \cdot Thickness$$

This relation places a strict upper bound on the prediction of *fresh weight* for any given *height*, *width*, and

thickness. However, because the box has space around the pad as well as the pad itself, this prediction

will consistently overestimate the actual volume (Figure 3).
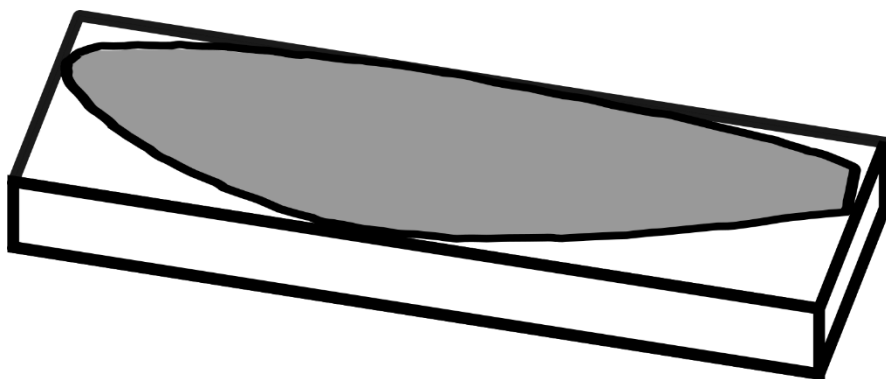
Figure 3. A schematic illustration of how the pad image relates to the rectangular "box" approximation of the pad's volume. This volume will always be larger than the actual pad volume but will never underestimate the pad volume.

To improve on this approximation, we had to use a more precise determination of pad area. This was done

exactly with manual area measurements, which were collected in ImageJ. The resulting "fitting box"

model, assuming a fairly uniform thickness, was presumed to provide a substantial improvement over the

"box" model (Figure 4).

EQ 6

$$FW \sim Area \cdot Thickness$$
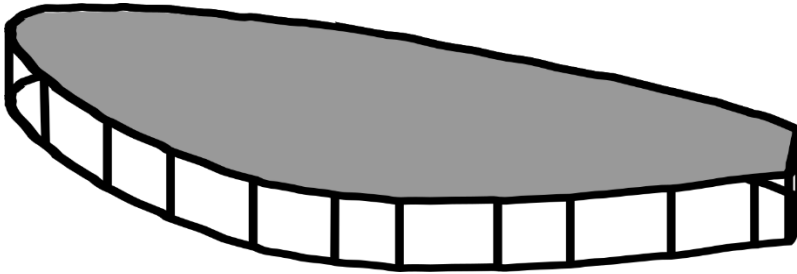
Figure 4. A schematic illustration of how the pad image relates to the "fitting box" approximation of the pad's volume. This volume will almost always be larger than the actual pad volume but may have error according to the precision of measurement of the pad's outline.

However, because measuring the area manually was relatively time consuming, another iteration of formulas was tested to determine the efficacy of an ellipse model for determining pad area (Figure 5).
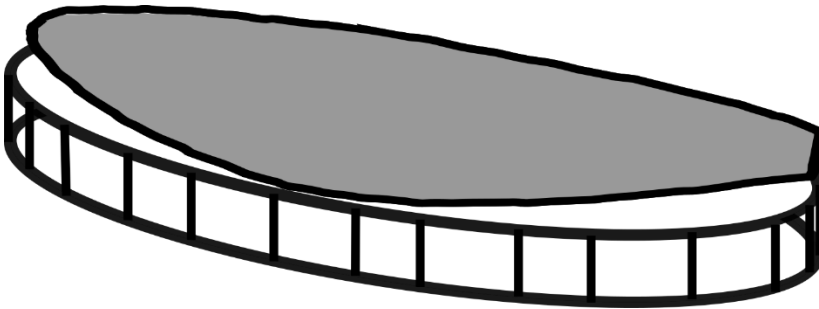
Figure 5. A schematic illustration of how the pad image relates to the "Ellipse" approximation of the pad's volume. This volume will be approximate to the pad volume but may over- or under-estimate the volume depending on the eccentricity of the pad.

These "Elliptical" models used the height and width of the pad to approximate the area of the pad according to the equation

EQ 7

Theoretical Elliptical Area

$$(A_{TE}) = \frac{\pi}{4}HW$$

And the equation for the "elliptical" model became

EQ 8

$$Fresh\ Weight \approx Elliptical\ Area \cdot Thickness$$

If area could be determined effectively with only two length measurements, then it would provide an improvement over both the "box" and "fitting box" models in terms of accuracy for the time and resources used.

The model sets described above will not fit perfectly to the pad as a direct volumetric method might – space might occur around the sides of the pad due to nonuniform pad thickness. To determine whether this extra space had a material effect on prediction accuracy, further investigation into the correlation between measures needed to be conducted.

Analysis of the initial five measures with Q/Q plots determined that assumptions of normality in the data were not supported. Hence, no assumptions of normality were made going forward in any of the analyses.

2.4 | Data Exploration

To determine the extent of variation between accession morphologies, violin plots of each of the five initial measures by accession were generated (Figure 6). This analysis was done to mathematically quantify the visually assessed differences in the ranges of the measurements among accessions (see Figure). These differences would determine the final formula's power to differentiate among accessions.

To determine the degree to which each measure varied with others in each accession, inter-measure correlations were created between each possible pair of the original five measures (Table 1 and Figure 7). This analysis was done to determine which terms were necessary to include or disclude from the final formula after testing, as determined by how much unique information they encoded in conjunction with other measures.

Whenever a model or set of models was being evaluated, the analysis took place at two scales – model performance fitted to individual accession data, and models fitted to the entire dataset at once (termed "composite data") without distinguishing among accessions. This analytical step was performed to determine whether the differences among accessions quantified in the statistical exploration step affected the fit quality of the model.

For each model that was developed, model statistics were extracted without also extracting model parameters. The goal of this analysis was to evaluate the comparative performance of the models by their form only. Without repeated trials, interpreting the value of the parameters would not be statistically meaningful – any number of confounding factors would muddle the interpretation of the differences between parameter values.

2.5 | Inclusion of Diameter

During the initial data exploration, diameter as a measure was intercorrelated well enough with other measures that the measure's inclusion in the formula would provide reasonable improvements to formula accuracy for the parsimony costs it incurred. Hence, diameter was tested in all three classes of formulas ("box", "fitting box", and "elliptical") going forward.

The modified formula for the "box" model is simply:

EQ 9

$$Fresh\ Weight \approx Height \cdot Width \cdot Thickness \cdot Diameter$$

And for the "fitting box" model, simply

EQ 10

$$Fresh\ Weight \approx Area \cdot Thickness \cdot Diameter$$

However, for the "elliptical" model, obtaining a diameter measure is less straightforward. The perimeter of the ellipse is required to calculate diameter, and the exact formula for the perimeter of an ellipse

requires a transcendental function (Sykora;, 2005). However, several approximations exist that are both

reasonably precise and compute efficiently with low margins of error.

To express the diameter of an ellipse given the cladode height and width, we can use:

EQ 6

Theoretical Elliptical Perimeter

$$(P_{TE}) = \frac{\pi}{4}(H + W)\frac{64 - 3h^2}{64 - 16h}$$

EQ 7

$$h = \left(\frac{H + W}{H - W}\right)^2$$

The approximation of a P_{TE} in EQ is taken from the Padé 2/1 (Hudson-Lipka) approximation (Hudson

and Lipka, 1917). This approximation was chosen because it was visually simple and computationally

inexpensive while also having adequate precision for the analysis – roughly 1 part in 900 inaccuracy

(Sykora;, 2005), which falls well below the rounding error in our study's initial measurements. Applying

the formula for the perimeter of a circle to the quantity P_{TE} creates a theoretical diameter measure (D_{TE}) as

needed. These approximations were substituted into the existing regression formula to create:

EQ 13

Fresh Weight ≈ Elliptical Area · Thickness · Elliptical Diameter

2.6 | Data Filtration

The data table containing the initial observations for all 14 accessions and 488 samples is included in

Appendix A1 for ease of viewing. During the initial statistical analyses, several significant outliers with

inconsistent proportions were found. Two categories of outliers were considered: realistically

proportioned cladodes that did not reflect general trends over the whole dataset, and "nonsense" cladodes

that could not exist as described (ex. excessive *diameter* relative to *width*, *width* higher than *height*). After

repairing the data from the original records and extensive cross-referencing, the cladode data were filtered

on the leverage they exerted on different inter-measure fits, excluding the relation of *thickness* with the

other three predictor measures, because outliers in those correlations were generally not outliers in other relations. The dataset filtered on leverage is referred to as the "3P/N" dataset, in reference to the formula used to determine the cutoff value of the threshold of a given plotted point (p for parameters, n for number of datapoints) (Science, 2022). This filtration not only consistently excluded nonsense and plausible outliers, but also was shown to have little effect on the general trends of each measure for each accession with *fresh weight*. However, all analyses were conducted in parallel on the unfiltered and filtered datasets, and all graphs not included in this thesis are available on request from the Github repository for this project (https://github.com/GabrielAngres31/CAI-Analysis-Script).

2.7 | Data Augmentation

To aid in further calculations and intercomparison with other models found in prior work, both the unfiltered and filtered datasets were augmented with several additional measures. Experimental area values for the "fitting box" models were taken using the imaging and measurement software ImageJ, using a ruler included in each photo as a reference for length. One area value was missing without the possibility of recovery at the time of analysis, so that data point was discarded for future area-related analyses. These area values are attached in Appendix A2. Measures for elliptical perimeter, area, and diameter were calculated and added for each pad. Lastly, the dry weight of each pad was approximated using constants of proportion derived from dry weighing and added to the dataset.

2.8 | Linear Modeling

To implement the models, the mathematical relations must be formatted as linear regression models. Thus, where c_0 and c_1 are constants of regression – with variable names condensed for brevity – the "box" model as a function of height, width, thickness, and diameter becomes:

EQ 14

$$FW \sim c_0 \cdot H \cdot W \cdot T \cdot D + c_1$$

Similarly, the "fitting box" model becomes:

EQ 15

$$FW \sim c_0 \cdot A \cdot T \cdot D + c_1$$

And the "elliptical" model becomes:

EQ 16

$$FW \sim c_0 \cdot Ell\_A \cdot T \cdot Ell\_D + c_1$$

The working hypothesis in this case is that all predictor terms, combined, provide the best overall model performance (accounting for the addition of terms on fit performance and model parsimony). Several models consisting of single or multiple variables in combination were set in a linear regression *vs.* fresh weight and evaluated over each accession separately and all accessions in aggregate, using two metrics to evaluate performance. The first metric was the adjusted coefficient of determination, or adjusted R^2 value, which augments a typical R^2 calculation by introducing a penalty for each additional term in the model, incorporating a measure of parsimony for model performance evaluation. The second measure, the Schwarz-Bayes criterion, is a strong measure that evaluates a model for parsimony rather than prediction accuracy – the smaller the value, the more parsimonious the model as compared to other formulations. This value is not an absolute value like adj. R^2 and is only used in comparison to values generated from other models on the same dataset, with the same independent variables. R Language, the statistical programming utility used throughout the analysis, has built-in functions to evaluate adj. R^2 and SBC metrics for linear model data. The formulas for these two metrics are shown in EQ and EQ .

EQ 17

$$Adj.R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

Where R^2 is the sample R-square for the model,
p is the number of predictors in the model,
and N is the total sample size for the model.

EQ 18

$$SBC = p ln(N) - 2 ln(\hat{L})$$

Where p is the number of predictors in the model,
N is the total sample size for the model,
and $\hat{L}$ is the maximized value of the likelihood function for the model of the mean of the standard normal

distribution, with the actual observed values represented as the mean with error term.

An important note: all linear regression analyses displayed here include interaction terms between each of the variables (as denoted by the * operator). Analyses were also generated without interaction terms, but as their performance was similar and only slightly worse, they were excluded from the main analysis and are available in the Appendix (A4)

2.9 | Methodology for Comparison to Prior Work

After the development of the formulas for fresh weight in our own research, we needed to determine how well formulas from other studies would perform on our dataset, and how our own formulas performed in comparison to their work. Hence, we conducted comparisons with the formulas derived in Tiznado-Hernandez et. al. 2010, Reis et. al. 2018, and Lucena et. al. 2021. Our analysis was not able to create a meaningful comparison with the results in Pinheiro et. al., (2015) due to substantial differences in modeling approaches. As Pinheiro et. al., 2015 had already developed linear and nonlinear equations for the surface area of their accessions of interest, the regression statistics those authors derived came from a dataset with two applied layers of regression (Pinheiro et al., 2015). Hence, applying the results of this analysis to the data of Pinheiro et. al., 2015 and *vice versa* would not be possible.

In Tiznado-Hernandez et. al., 2010, the authors determined the efficacy of approximating pad area with an ellipse – a technique that our research attempted to employ for the "elliptical" model we tested. Hence, we compared the model performances in their analysis and ours, using p-value comparisons, and testing fit efficacy over our entire input dataset and over individual accessions, both filtered and unfiltered (Table 2).

In Reis. et. al., 2018, the authors attempted to model cladode area (CA), biomass (FW), and dry weight (DW) of *Opuntia* cladodes by extrapolating from the width, length (synonymous with height), thickness, and the diameter of the neck of the cladode. The author's analysis included testing multiple

regression models for each parameter, with as few as five and as many as 17 models being tested for each

target statistic. Ultimately, three models were established for the target statistics, quoted directly with

fitted values as:

EQ 8

$$CA = 48.13 + 0.75(L \cdot W) \ [R\text{\textasciicircum}2 \ = \ 0.91]$$
$$FW = 36.91 + 0.64(L \cdot W \cdot T) \ [R\text{\textasciicircum}2 \ = \ 0.91]$$
$$DW = 8.49 + 0.49(W \cdot T \cdot D) \ [R\text{\textasciicircum}2 \ = \ 0.72]$$

These equations are linear regressions like the ones determined earlier to determine fresh weight. Three

analogous models based on our current dataset were compared to those in Reis et. al., 2018.

Lucena et. al., 2021 tested several formulae for the estimation of pad weight and pad area,

ultimately determined that a Power Law of the form

EQ 20

$$CA = c_1(HW)^{c2}$$

was best for determining area from height and width, whereas a gamma law of the form

EQ 21

$$FW = c_1HW + c_2T$$

was best for determining the fresh weight of a cladode from height, width, and thickness (Lucena et al.,

2021). Hence, we tested how these formulas worked on our own dataset, and how these formulas

compared in performance and parsimony to our three original models ("box", "fitting box", and

"elliptical").

2.10 | Figure Generation

Violin plots and heatmap plots were used to visualize statistical perspectives on the data that were

generated using a computer script in R (R Development Core Team, 2010). The script was designed in a

modular fashion to allow rapid custom generation of plots based on a choice of data (in the case of the

violin plots) or a choice of formulae to analyze over the given data (for the heatmaps). In total, ten violin

plots were generated: five for each of the initial measures of height, width, diameter, thickness, and fresh

weight; for each of the initial and filtered datasets. Additionally, 36 heatmaps were generated as follows. First, one adjusted R^2 heatmap was generated for measure intercomparison on the unfiltered dataset (supplemented by two three-dimensional scatterplots of the measures). Second, triplets of heatmaps comparing different suites of models (one heatmap for models without components of interaction (using the + operator), one model with components of interaction (using the * operator), and one heatmap with both sets at once) were generated for the adjusted R^2 and SBC measures, over both the "box" models and "fitting/elliptical" models, over the filtered and unfiltered datasets, making an additional 24 heatmaps. Lastly, heatmaps were generated to compare the results of Reis et. al., 2018 (three model sets on the unfiltered dataset in our analysis, with adjusted R^2) and Lucena et. al., 2021 (for the gamma law for fresh weight and the power law for area, with adjusted R^2 and SBC values, over the filtered and unfiltered datasets) resulting in an additional 11 heatmaps. Multiple, redundant summary plots were generated, and only a select few are presented in the results of this paper for conciseness.

All statistical analyses and figure/table generation was conducted in the RStudio terminal, and the generating program in its current state is attached in an appendix (A5). Initial single-measure analyses and inter-measure comparisons were conducted with several goals in mind: to determine the statistical properties of the data, to determine whether any measures correlated significantly with others or with *fresh weight*, and to find clear outliers in the data that might distort the results.

3 | Results

3.1 | Statistical Exploration of the Data

Analysis of the distribution of the five input measures established statistically significant differences in the ranges and tendencies of each measure over each accession (Figure 6). Each range had a unique distribution in terms of range, density, and distribution shape, with no general trends distinguishable between accessions. For instance, PARL accession 584 had a consistently wide distribution across all measures compared to other accessions, while 582 had a consistently narrower distribution. Accessions

242 and 246 had smaller rangers in thickness and fresh weight compared to other measures; 854 nearly

followed this trend exactly but had a wider distribution over thickness than 242 and 246. Altogether, there

were no discernable categories of pad morphology. All violin plots generated are included in Appendix

A3.
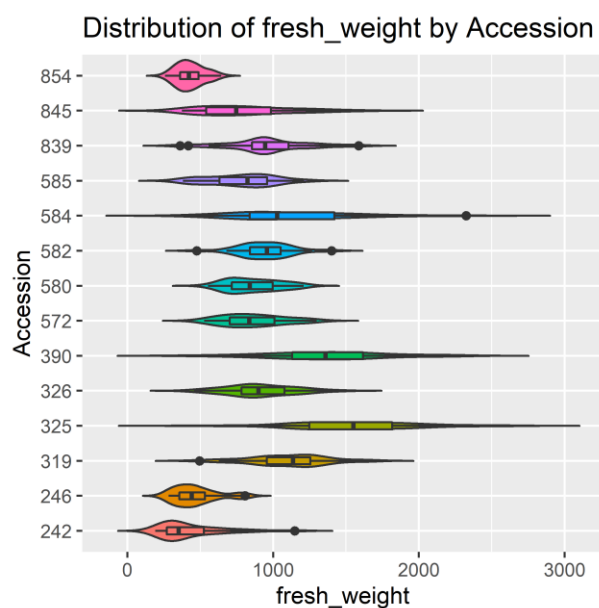
Figure 6. A violin plot of fresh weight value distributions for each PARL accession generated from the *unfiltered* dataset of the initial measures. Outliers (black dots) are highlighted. Interquartile ranges have been included in each violin shape as rectangles within the violin shapes.

3.2 | Measure Intercorrelations

Inter-measure correlations between the five input measures agreed with several intuitive judgements

about the relation of height, width, and thickness to fresh weight. Fresh weight correlated adequately well

with width (mean 0.60, range 0.44 to 0.76, composite 0.54 adjusted R^2) and height (mean 0.46, range

0.15 to 0.77, composite 0.37 adjusted R^2) and poorly with thickness (mean 0.18, range 0.01-0.43,

composite 0.35 adjusted R^2). In addition, thickness did not correlate well with width (mean 0.02, range

0.03 to 0.19, composite 0.04 adjusted R^2), height (mean 0.02, range -0.03 to 0.11, composite 0.06

adjusted R^2), or diameter (mean 0.02, range -0.03 to 0.13, composite 0.08 adjusted R^2) (Figure 7,

Table 1). However, diameter had significant correlations with height (mean 0.61, range 0.05 to 0.94,

composite 0.57 adjusted R^2), width (mean 0.39, range -0.03 to 0.69, composite 0.18 adjusted R^2) and

fresh weight (mean 0.54, range 0.00 to 0.87, composite 0.49 adjusted R^2). This observation was used as

the basis for including diameter in the basic three formulas developed in this research (Figure 5, Table 1).

Accessions 319, 580, and 854 consistently performed more poorly on all inter-measure correlations than

other accessions, while 845, 326, and 242 consistently performed better than other accessions.
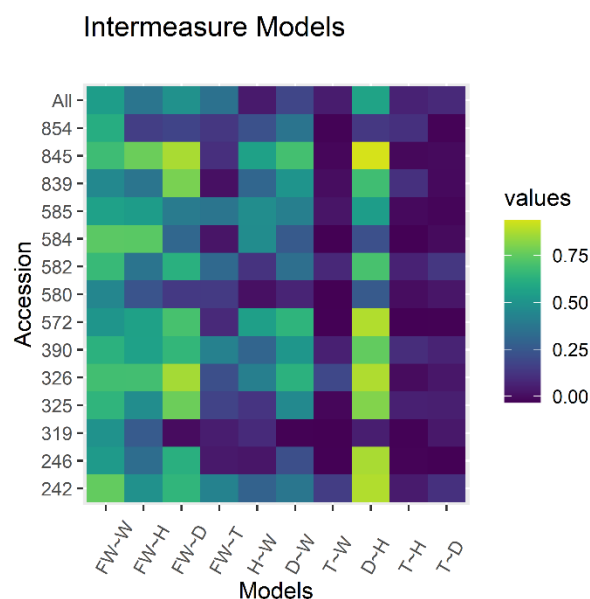
Figure 7. A heatmap plot of various inter-measure comparisons in different accessions, on a linear fit model, using an unfiltered dataset, evaluated using an R^2 value. Lighter values indicate a higher adj. R^2 using the viridis color palette. The abbreviations FW, W, H, D, and T stand for fresh weight, width, height, diameter, and thickness, respectively.

Table 1. A table of various inter-measure comparisons in different accessions, on a linear fit model, using an unfiltered dataset. The abbreviations FW, W, H, D, and T stand for fresh weight, width, height, diameter, and thickness, respectively. Notable relationships are emphasized in bold.

Access-ion	FW~W	FW~H	FW~D	FW~T	H~W	D~W	T~W	D~H	T~H	T~D
All	**0.537**	**0.369**	**0.486**	0.350	0.035	0.184	0.042	**0.567**	0.062	0.084
854	**0.611**	**0.154**	**0.175**	0.129	0.220	0.362	-0.028	**0.133**	0.106	-0.027
845	**0.681**	**0.770**	**0.870**	0.105	0.562	0.695	-0.019	**0.938**	-0.013	-0.009
839	**0.452**	**0.368**	**0.794**	0.010	0.307	0.505	-0.001	**0.685**	0.107	-0.010
585	**0.553**	**0.533**	**0.392**	0.368	0.469	0.407	0.021	**0.539**	-0.011	-0.021
584	**0.739**	**0.736**	**0.310**	0.021	0.465	0.257	-0.033	**0.217**	-0.032	-0.007
582	**0.665**	**0.363**	**0.625**	0.319	0.117	0.340	0.080	**0.703**	0.063	0.131
580	**0.444**	**0.234**	**0.139**	0.141	0.007	0.071	-0.033	**0.255**	0.001	0.023
572	**0.508**	**0.561**	**0.701**	0.087	0.547	0.641	-0.032	**0.883**	-0.031	-0.031

390	**0.630**	**0.555**	**0.651**	0.417	0.297	0.512	0.059	**0.754**	0.097	0.064
326	**0.689**	**0.690**	**0.863**	0.214	0.407	0.630	0.191	**0.878**	-0.004	0.027
325	**0.642**	**0.469**	**0.771**	0.171	0.124	0.455	-0.020	**0.809**	0.059	0.051
319	**0.490**	**0.261**	**-0.001**	0.047	0.089	-0.031	-0.032	**0.050**	-0.030	0.031
246	**0.528**	**0.332**	**0.623**	0.033	0.023	0.215	-0.031	**0.870**	-0.027	-0.032
242	**0.756**	**0.484**	**0.644**	0.427	0.288	0.375	0.149	**0.884**	0.038	0.107

To better visualize the relationships between the initial measures, 3D scatterplots were made to plot the relationship between height, width, and thickness, and between thickness, diameter, and fresh weight (Figure 8a,b, 9a,b). Confirming the results of the violin plots, each accession appears to occupy its own distinct range of values (Figure 8a,b), while nonetheless conforming to a central axis of cactus morphology (for instance, the range of fresh weight, thickness, and diameter values are largely confined to a plane) (Figure 9a,b).
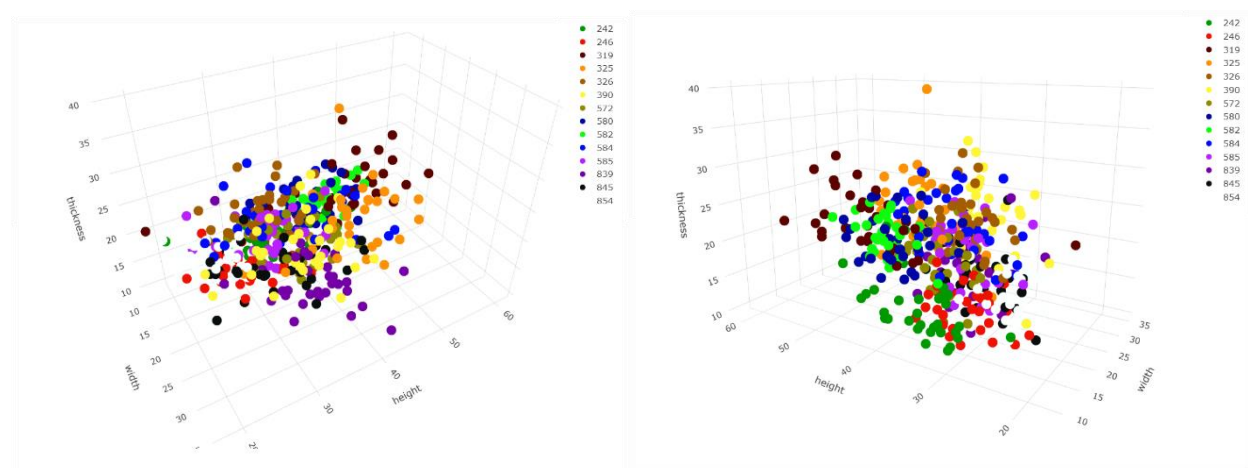
Figure 8. Two perspectives (A, left and B, right) of a 3D scatter plot of cladode height, width, and thickness colored by accession. All accessions are color-coded according to the key in the upper right corner on each graph.
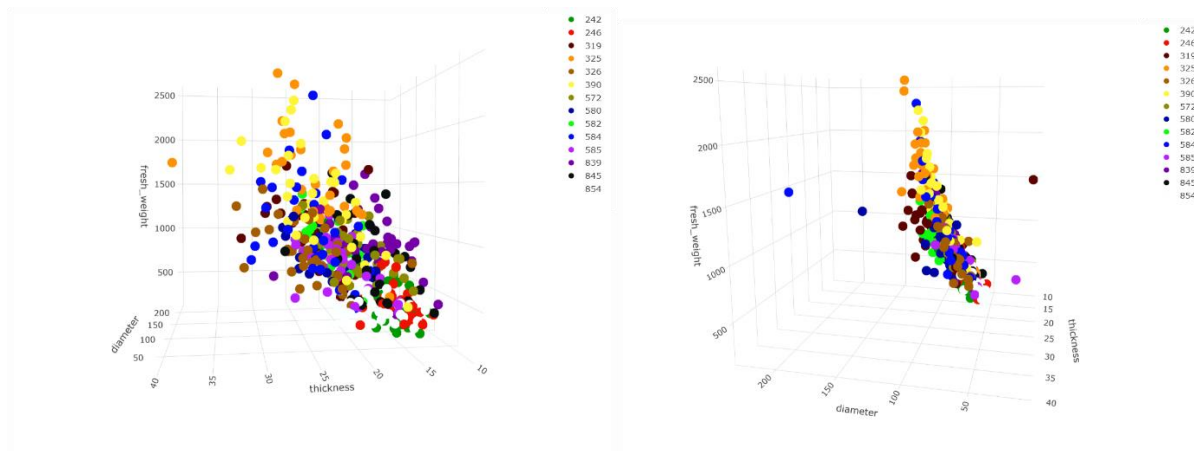
Figure 9. Two perspectives (A, left and B, right) of a 3D scatter plot of cladode thickness, diameter, and fresh weight. All accessions are color-coded according to the key in the upper right corner on each graph.

3.3 | "Box" Model Testing

To verify the efficacy of the "box" models, all possible sub-combinations of the models were tested: four single-measure comparisons, six two-measure comparisons, four three-measure comparisons, and the full model with four variables. Out of *width, height, diameter,* and *thickness,* only the first three measures provided meaningful contributions to accession-specific models by themselves, while thickness did not (Figure 10). Height and width correlated well with each other and boosted model performance relative to models that did not include both height and width; of the two, however, width provided the greater contribution to model fits (Table 1). Meanwhile, thickness, while a poor predictor of fresh weight by itself, improved accuracy of fit in combination with other measures (Figure 10). Omitting height or both height and width produced models with close-to-optimal parsimony, and for all but one accession, diameter was part of the most parsimonious model (Figure 11). Alternatively, the imprecision could result from the differences in measure intercorrelation between each accession (see Figure 7). This explanation is supported by several inconsistencies in the general trends in correlation to fresh weight. For instance, accession 585 had a significantly different performance with relating thickness to fresh weight and had a relatively smaller range of fit values over the models tested, while accession 580 had consistently poor performance. Several models (diameter, width*diameter, and height*diameter (D, W*D, H*D)), returned

relatively poor or good performance as a group. Specifically, if an accession performed well on one of those models, it was likely to perform well on the other two. This trend was also connected with the accession's performance on the "full" model (all variables in combination). This confirmed that the diameter measure is an important contributor to the full model's performance.
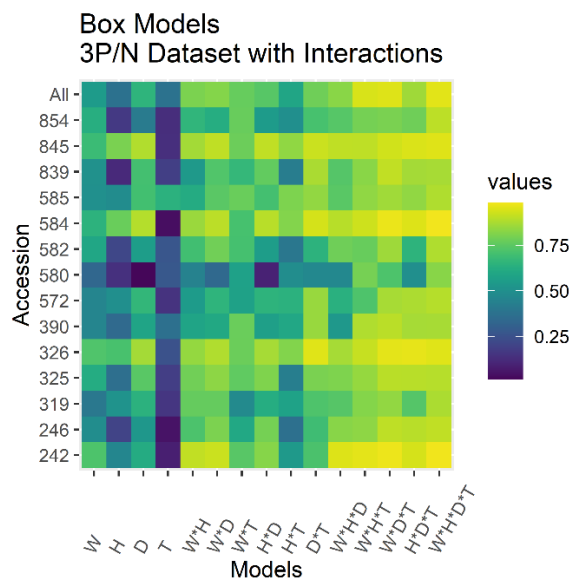
Figure 10. A heatmap plot of all possible combinations of the four base measures for the "box" model, evaluated using an adjusted R^2 value. Yellow (lighter color) indicates an R^2 value of 1, while dark blue (darker color) indicates a value of 0 or below using the viridis color palette.

Figure 11. A heatmap plot of all possible combinations of the four base measures for a "box" model, evaluated using a normalized SBC criterion value. Interaction terms are included, as indicated by the "*" operator. The lighter color indicates the lowest and most optimal SBC value for the row, while the darker color indicates the highest value and least parsimonious model using the viridis color palette.

Generally, the "full" model returned the best results in terms of adj. R^2 value (Figure 10) and parsimony (Figure 11). However, for individual accessions, the full model was not the most parsimonious (Figure 11). This discrepancy between the SBC table and adjusted R^2 table was likely due to the differences in morphology among accessions.

3.4 | "Fitting Box" and "Elliptical" Model Testing

To verify the efficacy of the "fitting box" and "elliptical" models, models with area values were tested with and without the inclusion of thickness, then with and without the inclusion of pad diameter/ellipse "diameter" as needed (Figure 12, 13). Generally, performance for these models was comparable to the "box" models. Contrary to intuition, however, the ellipse equations performed comparably or better than the models that used manual area measurements. The range of performance of manual area models between the best and worst performing accessions was 0.81, 0.40, and 0.21 points of adjusted R^2 for area, area*thickness, and area*thickness*diameter, respectively while the range of performance of the analogous equations for the ellipse models were 0.46, 0.21, and 0.24. Hence, the performance of models

that used elliptical area approximations was more homogenous <<than what?>>. Over individual

accessions, models that used manual area measurements outperformed models that used elliptical

approximations by 0.015 to 0.085, and (in the case of accession 572) by as much as 0.16 points of

adjusted R^2 difference. Despite this, using the elliptical model with diameter area measure yielded an

adj. R^2 of 0.94 over *all* accessions treated as one group, with less than 0.01 difference in fit compared to

the manual area model.

From a parsimony perspective, diameter is a detriment to model performance at the accession-

specific level and a benefit to model performance at the level of the whole dataset. Area measures by

themselves, without the inclusion of thickness, were too heterogenous to be consistently parsimonious

even on the general level (Figure 13). In addition, when comparing theoretical area and manual area

models, composite models achieved comparable performance within 0.02 points of adjusted R^2 value.
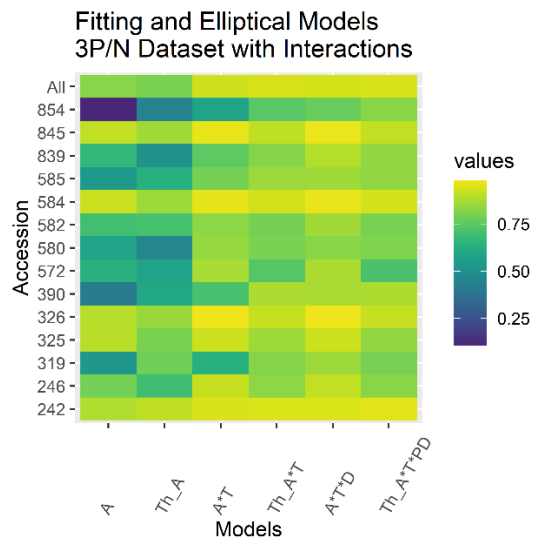
Figure 12. A heatmap plot of the "fitting box" and "elliptical" approximation models being evaluated, with subcomponents less area and diameter for comparison, evaluated using an adjusted R^2 value. Yellow (lighter color) indicates an R^2 value of 1, while dark blue (darker color) indicates a value of 0 or below using the viridis color palette.
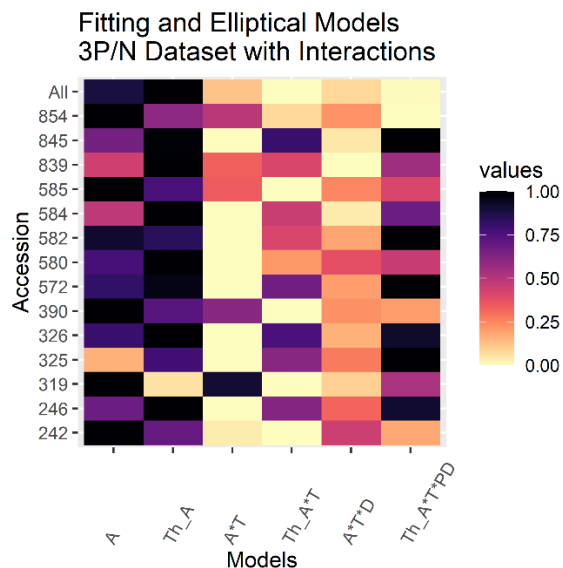
Figure 13. A heatmap plot of the "fitting box" and "elliptical" approximation models being evaluated, with subcomponents less area and diameter for comparison, evaluated using a normalized SBC crit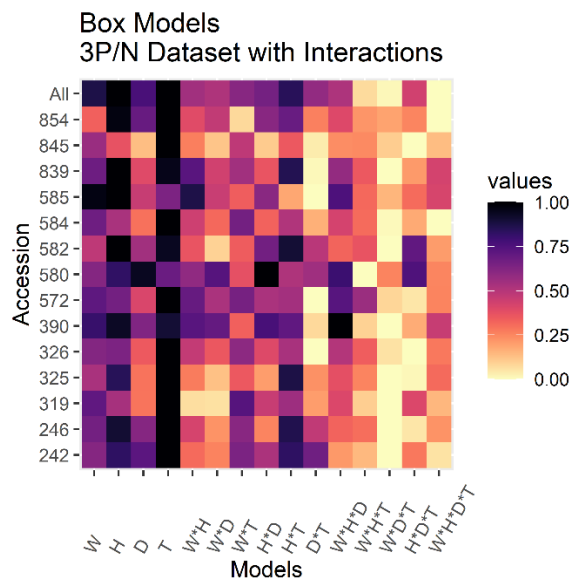erion value. The lighter color indicates the lowest and most optimal SBC value for the row, while the darker color indicates the highest value and least parsimonious model using the viridis color palette.

3.5 | Comparisons to Previous Analyses

In comparison with Tiznado-Hernandez et. al., 2010, we established a comparable standard of accuracy of fit between manually measured and computed area. Qualitatively, the authors corroborate the idea that the ellipse model is appropriate for modeling cladode area. In two of the *Opuntia* varieties that the authors analyzed ("Atlixco" and "Copena F-1"), R^2 values of 0.94 and 0.96, respectively, were obtained for the correlation between theoretical and experimental area. However, the authors did not report R^2 values for the other two varieties tested (including the "Milpa Alpa" and "Jalpa" varieties). Our analyses showed that it was possible to achieve similar R^2 values to Tiznado-Hernandez's 2010 analysis with a smaller sample size (as with 242, 584, 585, and 854); however, many accessions did not reach the same performance as Tizando-Hernandez's 2010 work. This difference in performance may be in some part due to the lowering effect of the adjusted R^2 correction. However, given that nearly every accession in our analysis met or fell below the $p < 0.0001$ threshold that the authors achieved, the difference is more likely due to inter-accessional variation (Tiznado-Hernández et al., 2010) (Table 2).

Table 2. Comparison of Experimental to Elliptical Area in Tiznado-Hernandez, 2010 *vs*. an analysis in our dataset, both filtered and unfiltered. Adjusted R^2 values from this analysis, as well as collected P-Values, for the correlation between experimental and elliptical area were compared to the analyses of Tiznado-Hernandez et. al. (2010).

Accession	Adj. R^2		P-Value	
Tiznado-Hernandez et. al.	Experimental vs. Theoretical Area			
Copena F-1	0.9637		0.0001	
Atlixco	0.945		0.0001	
Current Analysis	Unfiltered	3P/N	Unfiltered	3P/N
All	0.8824	0.91	<0.0001	<0.0001
242	0.9721	0.9777	<0.0001	<0.0001
246	0.862	0.8122	<0.0001	<0.0001
319	0.2789	0.5177	0.0011	<0.0001
325	0.9074	0.8964	<0.0001	<0.0001
326	0.9225	0.924	<0.0001	<0.0001
390	0.814	0.7086	<0.0001	<0.0001
572	0.7886	0.7291	<0.0001	<0.0001
580	0.835	0.7635	<0.0001	<0.0001
582	0.8248	0.7545	<0.0001	<0.0001
584	0.9607	0.9545	<0.0001	<0.0001
585	0.5842	0.5383	<0.0001	<0.0001
839	0.7474	0.6308	<0.0001	<0.0001
845	0.9408	0.9404	<0.0001	<0.0001
854	0.2916	0.2897	0.0008	0.0011

In comparison with Reis et. al., 2018, testing over regressions to area, fresh weight, and dry weight had mixed results in terms of comparative performance. For the determination of pad area, the general performance over the composite dataset in our analysis was comparable to the author's results. With three different example formulas (rectangular area, elliptical area, and elliptical area with diameter), the average performance over specific accessions was 0.79, 0.77, and 0.81 adjusted R^2, respectively. Excluding 319 and 854 (the consistently low values) raised the values up to 0.86, 0.85, and 0.88 – comparable to the value of 0.88 obtained for all composite models, but still less (though comparable) than the 0.91 R^2 that Reis achieved (Figure 14). Meanwhile, our analysis produced an improvement in measuring fresh weight. The models tested for dry weight were the box models with and without diameter, and the full area elliptical model, and reached an average performance of 0.84, 0.90, and 0.85 adjusted R^2, respectively. While this does not exceed Reis et. al., 2018's value of 0.91 R^2, the

performance of each model over the composite dataset is superior at 0.94, 0.96, and 0.94, respectively

(Figure 15). Lastly, our analysis found higher general and accession-specific performance in measuring

dry weight, than Reis et. al., 2018. The authors obtained an R^2 value of 0.72 with their formula using

width, thickness, and pad neck diameter (Reis et al., 2018). Meanwhile, our analysis tested dry weight as

a function of width*thickness, width*thickness*pad diameter, and of elliptical area*thickness*pad

diameter, obtaining accession-wide averages of 0.72, 0.86, and 0.87 adjusted R^2, respectively, while the

composite data returned values of 0.80, 0.88, and 0.87 adjusted R^2, respectively (Figure 16). No

parsimony analyses were run because we did not have access to their data and did not take a neck

diameter measurement. The goal of this comparison was solely to understand which models of the same

quantity would return a better performance on the same data.
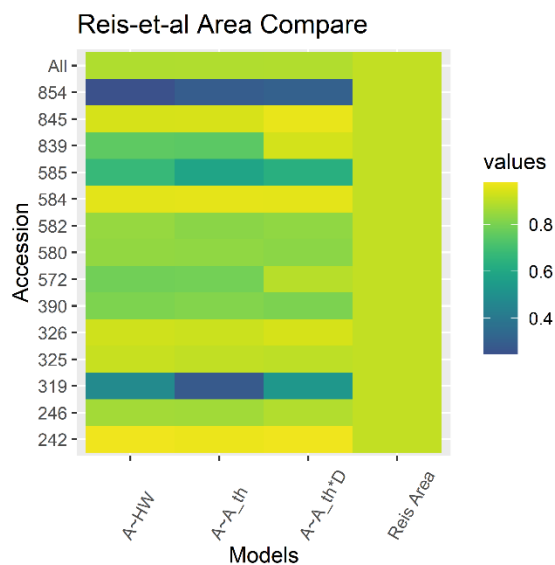
Figure 14. Comparison of regression effectiveness between the model for cladode area in Reis et. al., 2018, and three models explored in this analysis. Yellow (lighter color) indicates an R^2 value of 1, while greener colors indicate lower values using the viridis color palette. The abbreviations A, H, W, A_th, and D refer to area, height, width, theoretical area (elliptical area), and diameter, respectively.
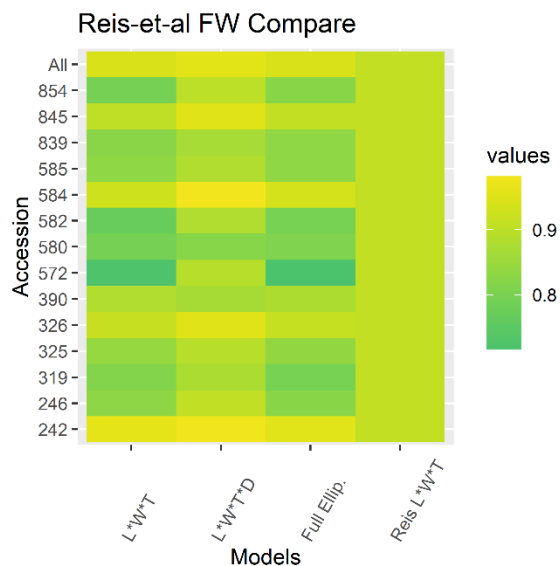
Figure 15. Comparison of regression effectiveness between the model for cladode fresh weight in Reis et. al., 2018, and three models explored in this analysis. Yellow (lighter color) indicates an R^2 value of 1, while greener colors indicate lower values using the viridis color palette. The abbreviations L, W, T, and D refer to length (height), width, thickness, and diameter, respectively. Full Ellip refers to the elliptical model tested earlier in this analysis (elliptical area * elliptical diameter * thickness).
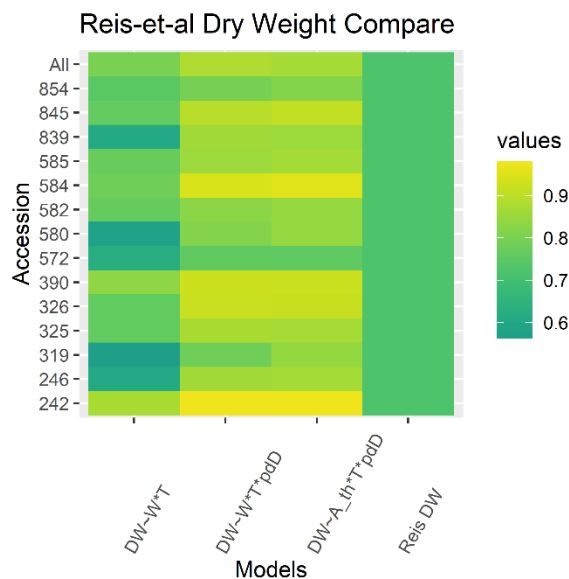
Figure 16. Heatmap plot of comparison of regression effectiveness between the model for cladode dry weight in Reis et. al., 2018, and three possible equivalents derived in this analysis. Yellow (lighter color) indicates an R^2 value of 1 while greener colors indicate lower values using the viridis color palette. The abbreviations DW, W, T, pdD, and A_th refer to dry weight, width, thickness, Pade-derived diamete (elliptical diameter), and theoretical area (elliptical area).

For the comparison to Lucena et. al., 2021's formulae, we found that their formulae showed a strong performance on some individual accessions, but largely did not outperform our models on the same data. We replicated their formulas on our dataset to determine their efficacy on our data and compare those formulas to our box, fitting box, and elliptical formulae. Over this project's dataset, the author's Power Law to determine area had good performance over single accessions (12 accessions Adj. R^2 = 0.75-0.97, 2 accessions Adj. R^2 < 0.40) and Adj. R^2 = 0.88 composite (Figure 17), while also performing better than either height- or width-based power laws by themselves (Figures 17, 18).
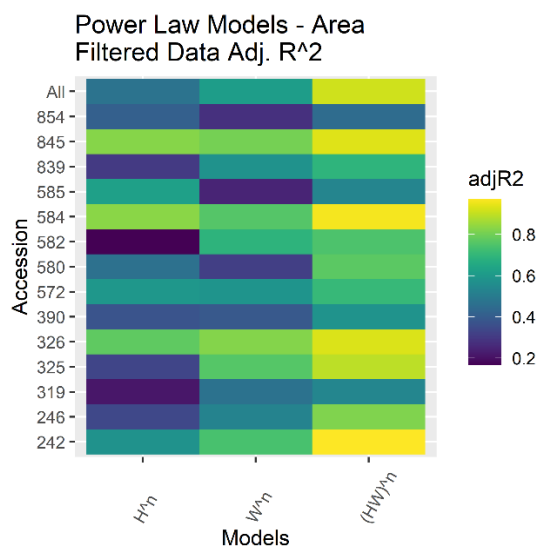
Figure 17. A heatmap plot of the Power law relationship between cladode height and width vs. pad area, explored in Lucena et. al., 2021, along with single-measure subcomponents of the model, evaluated on an Adjusted R^2 parameter. Yellow (lighter color) indicates an R^2 value of 1, while greener colors indicate lower values using the viridis color palette. The abbreviations H and W refer to height and width, respectively.
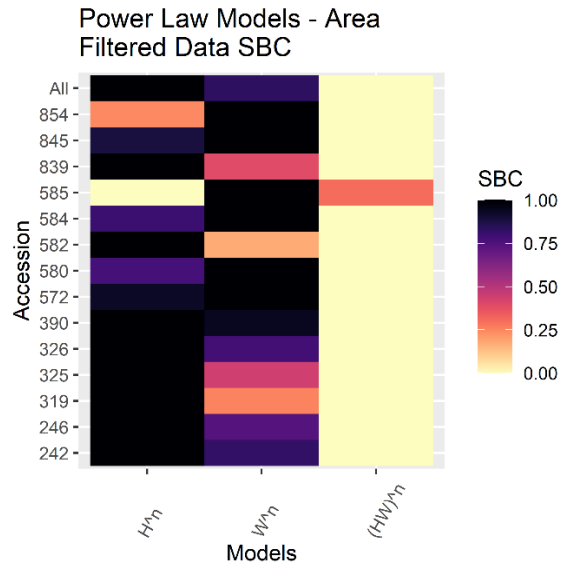
Figure 18. A heatmap plot of the Power law relationship between cladode height and width vs. pad area, explored in Lucena et. al., 2021, along with single-measure subcomponents of the model, evaluated on a normalized SBC criterion value. The lighter color indicates the lowest and most optimal SBC value for the row, while the darker color indicates the highest value and least parsimonious model using the viridis color palette. The abbreviations H and W refer to height and width, respectively.

The fresh weight gamma law models that Lucena et. al., 2021 developed had a high performance on our dataset, both over individual accessions (Adj. R^2 = 0.73-0.93, average performance 0.84) and in general (Adj. R^2 = 0.93) (Figure 19), while also having an unequivocal parsimony advantage over its subcomponents (Figure 20).
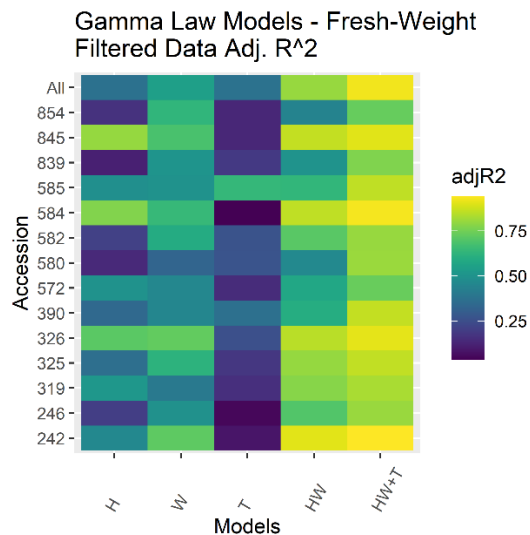
Figure 19. A heatmap plot of the Gamma law relationship between cladode height, width, and thickness vs. pad fresh weight, explored in Lucena et. al., 2021, along with single-measure subcomponents of the model, evaluated on an Adjusted R^2 parameter. Yellow (lighter color) indicates an R^2 value of 1, while greener colors indicate lower values using the viridis color palette. The abbreviations "H", "W", and "T" refer to "height", "width", and "thickness", respectively.
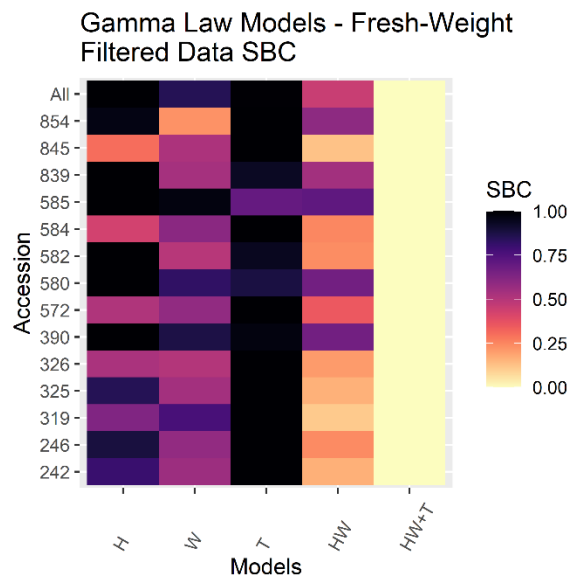
Figure 20. A heatmap plot of the Gamma law relationship between cladode height, width, and thickness vs. pad fresh weight, explored in Lucena et. al., 2021, along with single-measure subcomponents of the model, evaluated on a normalized SBC criterion value. The lighter color indicates the lowest and most optimal SBC value for the row, while the darker color indicates the highest value and least parsimonious model using the viridis color palette. The abbreviations H, W, and T refer to height, width, and thickness, respectively.

However, the Gamma model for fresh weight, as compared to the best-performing models from the three

model schemes developed in this project, consistently underperformed and was in many cases (including

in the whole-dataset analysis) the worst-performing model out of the four being compared (Table 3).

Table 3. A list of adjusted R^2 values for the top performing models from the three modeling branches in this report *vs*. the Gamma model for fresh weight developed in Lucena et. al., 2021. The box model is fresh weight as a function of height, width, thickness, and diameter with interactions; the fitted box model is fresh weight as a function of experimental area, thickness, and diameter with interactions; the elliptical model is fresh weight as a function of elliptical area, thickness, and elliptical diameter. The highest value in each row is bolded, and the lowest value in each row is italicized and underlined.

Accession	Box	Fitted Box	Elliptical	Gamma
All	**0.956524**	0.9355	0.94276	*0.925676*
242	**0.979809**	0.947384	0.955653	*0.94225*
246	0.906292	**0.908714**	0.823303	*0.804128*
319	**0.875142**	0.855113	*0.796175*	0.827164
325	0.891248	**0.923772**	*0.838654*	0.85881
326	0.953934	**0.977578**	0.914072	*0.907515*
390	0.868547	0.873163	**0.87794**	*0.861747*
572	**0.891075**	0.87921	*0.716373*	0.733647
580	0.821741	**0.822703**	0.807564	*0.806004*
582	**0.881471**	0.860322	*0.79663*	0.802007
584	**0.981808**	0.967965	0.934469	*0.929198*
585	**0.881648**	0.857918	*0.834055*	0.856126
839	0.869618	**0.889326**	0.834365	*0.769069*
845	0.952409	0.971091	**0.909154**	*0.901636*
854	**0.900627**	0.769108	0.823143	*0.731306*

4 | Discussion

4.1 | Insights into Differences Among Accessions

In the initial statistical explorations of the data, the variance in measure distribution among accessions

agreed with biological intuition of the factors governing cladode shapes. In nature, the phenotypic

presentations of organisms often occupy a spectrum. For example, the leaf shape of *Arabidopsis thaliana*

and its variations over hundreds of accessions can be attributed to multiple quantitative trait loci

(Vaughan, 2015). Leaf shape (or the shape of other plants parts) as a multifactorial genetic phenomenon can extend to cladode morphology. This explanation is consistent with the variety exhibited both in the strength or weakness of the inter-measure correlations for each accession and in how the distributions of different measures varied with each other.

Throughout this analysis, models generated on composite data over accession-specific models have produced a consistent advantage. Regardless of which model was being tested at the time, the average performance over individual accessions often fell well below the composite model fit. This demonstrates that cactus pear accessions adhere to a central axis of cladode morphology.

4.2 | Predicted vs. Actual Model Performance

The relative performance of the box, fitting box, and ellipse models for pad fresh weight did not match the expectations dictated by mathematical intuition. As stated before, the 3D models represented by each formula have a different tendency to match the shape of the actual cactus pad. A box with the same face area as the pad (the fitting box model) would be expected to have the greatest fit, due to the minimum of empty space around the sides, followed by the ellipse and box models. However, the experimental results of the modeling tell a different story. The fitting box model was the worst model, and the box and ellipse models were superior qualitatively and quantitatively. Over individual accessions, the box model was the best model for seven accessions, the second-best for six accessions, and the third-best for only one accession. Meanwhile, the ellipse model was the best-performing model for six accessions, the second-best for six accessions, and the third-best for two accessions. Furthermore, the fitting box model was only optimal for one accession, second-best for two accessions, and third-best for eleven accessions. Numerically, the average ordinal rankings of the models were 1.57, 2.71, and 1.71 for the box, fitting box, and ellipse models, respectively. Over the composite data, the ranking from best to worst is the box, elliptical, and fitted box models. However, performance was more heterogenous from the actual adjusted

R^2 values. Performance over individual accessions decreased from the box models (0.822-0.982 adj. R^2, mean 0.904, range 0.160) to the fitted box models (0.769-0.978 adj. R^2, mean 0.893, range 0.208) to the ellipse models (0.716-0.956 adj. R^2 mean 0.847, range 0.239). However, over the composite dataset, the performance values were 0.957, 0.936, and 0.943 for the box, fitting box, and ellipse models, respectively. Although the model performances were comparable, the box and ellipse models had an advantage: a costly area calculation needed for the "itting box model was unnecessary for the other two models. Hence, we propose the following recommendations for cactus modeling efforts going forward:

- Use as large a sample size as is possible, because this will minimize the effects of heterogeneity among and within accessions on the model quality.

- Employ the box model if a means is available to measure the perimeter of the pad and derive a diameter measure accurately and efficiently.

- If a diameter measure cannot be feasibly derived, then the elliptical model will produce a comparable if slightly inferior performance.

4.3 | Insights from Prior Work Comparisons

The comparison of our work with that of Tiznado-Hernandez et. al., 2010 underscores the significance of quantifying variation between accessions. Even with a much smaller sample size over individual accessions (32 in our analysis vs. 80 in the author's analysis) (Tiznado-Hernández et al., 2010), several accessions performed comparably or above the values obtained in the author's analysis, and yet, for the same p-values, many of our accessions underperformed. Taking into consideration the consistently superior performance of composite models, we conclude that sample size and the degree of pad shape variation within accessions are essential considerations to evaluating an indirect model's performance.

The comparison of our work with that of Reis et. al., 2018, highlights the importance of evaluating the contribution of each component to the overall performance of an indirect model. That analysis established cactus pad base diameter as part of a formula for dry weight. However, when neck

diameter was used as part of a similar modeling effort in an earlier project by Reis, Ribeiro, and Gazarini in 2014 (Reis et al., 2014), neck diameter did not appear as part of an optimal modeling formula over the seven best performing accessions in their analysis. Neither analysis explicitly stated the rationale for including neck diameter in their models other than that the model was employed by a previous study or highlights this difference in performance in their earlier efforts. Without such comparisons, or any further investigations into why models do well or poorly on datasets, substantive contributions to indirect modeling in *Opuntia* cannot be produced.

The comparison of our work with that of Lucena et. al., 2021, indicated that different modeling principles may govern models over single accession *versus* models over composite data. The range of possible values for each measure is generally narrower for individual accessions than a composite dataset. Consequently, models that can generalize effectively between varying accessional morphologies will gain less performance when fitting on an individual accession as compared to a model that is more sensitive to accession-specific morphological characteristics. However, although different rules govern single- and composite-accession modeling, both should be investigated as approaches to deriving accurate and reliable models.

4.4 | Future Directions in Indirect Model Formulation

The extent to which the modeling efforts developed in our research can extend to other crops depends upon multiple external factors. The specific formulae developed here are unlikely to extend to non-*Opuntia* crops, as different plant part shapes will necessarily require different model geometries. However, the general approach of first-principles geometric modeling is suitable for any crop that has an easily computable volume approximation.

Further development of indirect models for cladode fresh weight in *Opuntia* requires specific research into the variations between accessions, both in obtaining data and in formulating equations. Physically, the demonstrable morphological differences that exist among *Opuntia* accessions may

introduce errors in measurement. For example, a thin pad that is misshapen will be more difficult to measure than a broader, thicker, and more rigid pad. In addition, biochemical factors beyond water content, such as nutrient sufficiency/deficiency states among accessions, need to be evaluated for any relevance to density and volume approximation considerations.

4.5 | Future Directions for Indirect Modeling Platforms

Sufficiently accurate general formulae for use on composite data creates a direct pathway towards large-scale trials based on indirect measurement. With easier *in situ* measurement in the field, the quantity of pads that can be efficiently sampled at one time will increase significantly. Researchers employing camera imaging of pads in the field supplemented by manual thickness measurements will be able to efficiently sample from a range of physically accessible pads. Farming productivity analyses that are primarily concerned with fresh, one-year-old cladodes at the ends of branches are well-suited to the use of such sampling methods.

In this thesis, we demonstrated that accurate prediction over a heterogenous dataset was possible with relatively crude one-dimensional measurements. However, the use of three-dimensional modeled volumes provides another milestone for indirect modeling efforts in *Opuntia*. Where one-dimensional measurements can simulate rough approximations of a cladode's actual volume, modern 3D object scanning technology can obtain (for the purposes of field measurements) arbitrarily precise maps of the cactus pads with automatically computed volumes. Light detection and ranging (LIDAR), a technology originally developed for terrain and forest surveying, is gaining traction as a proposed complement to 2D image analysis in agricultural analysis (Lin, 2015). Specifically, a measurement pipeline was developed for determining the surface area of maize and sorghum leaves (Thapa et al., 2018). The authors utilized a LIDAR scanning machine rotated around the sample plant with a full 360-degree range, then compiled the resulting LIDAR time-of-flight data into a point-cloud map. From there, standard 3D processing steps accurately rendered the leaves, obtaining a greater than >0.91 R^2 correlation for individual leaves and >0.95 R^2 for total plant leaf area as compared to manual measurements (Thapa et al., 2018). A similar

approach could, in theory, be implemented in *Opuntia* cladode measurement. The advantages of using LIDAR, besides the obvious benefit to precision, would be that cactus pads are generally more uniform in shape than the thin, non-succulent leaf structures of maize and sorghum and therefore easier for a point-cloud model to parse. In addition, extensive 3D modeling efforts in *Opuntia* that quantify the full range of accession-specific morphological variation would make accession-specific indirect models fully viable. However, several obstacles must be overcome before LIDAR modeling approaches are feasible. The dense arrangement of cladodes on sufficiently large, established trees creates uncertainty in the assumption that all points on the cactus tree will be accessible from a given range of laser beam angles and positions. In addition, the infrastructure will require full 360-degree access to each tree, requiring greater spacing between individual plants and increasing land use. These challenges restrict the current feasibility of LIDAR modeling to smaller or more portable trees.

5 | Conclusions

In this research, several sets of models, each with their own mathematical rationale and construction, were evaluated for their predictive quality over individual *Opuntia* accessions and over a composite dataset. The results showed that these models, and by extension the prediction of a pad's weight from its volume (or area) and eccentricity, have substantial potential for further use of cladode measurement data to estimate biomass productivity. Furthermore, the variable performances of these models at different levels were used to establish qualitative bounds on a given model's ability to generalize across accessions. The results of measure analysis and multiple composite modeling efforts demonstrated that a central axis of cactus morphology exists, and that composite modeling should be investigated alongside accession-specific models.

Works Cited

Adeel Z, Safriel U, Niemeijer D, White R (2005) Ecosystems and human well-being: desertification synthesis. World Resources Institute (WRI)

Agriculture Victoria (2022) A brief guide to estimating crop yields. **Access 2022:** https://agriculture.vic.gov.au/crops-and-horticulture/grains-pulses-and-cereals/crop-production/general-agronomy/a-brief-guide-to-estimating-crop-yields

Borland AM, Hartwell J, Weston DJ, Schlauch KA, Tschaplinski TJ, Tuskan GA, Yang X, Cushman JC (2009) Engineering crassulacean acid metabolism to improve water-use efficiency. doi: 10.1016/j.tplants.2014.01.006

Colaizzi PD, Evett SR, Brauer DK, Howell TA, Tolk JA, Copeland KS (2017) Allometric method to estimate leaf area index for row crops. Agronomy Journal **109:** 883-894, doi: 10.2134/agronj2016.11.0665

Davis SC, Simpson J, Gil-Vega KdC, Niechayev NA, Tongerlo Ev, Castano NH, Dever LV, Búrquez A (2019) Undervalued potential of crassulacean acid metabolism for current and future agricultural production. Journal of Experimental Botany **70:** 6521-6537, doi: 10.1093/jxb/erz223

Hijmans RJ (2003) The effect of climate change on global potato production. American Journal of Potato Research **80:** 271-279, doi: 10.1007/BF02855363

Huang J, Zhang G, Zhang Y, Guan X, Wei Y, Guo R (2020) Global desertification vulnerability to climate change and human activities. Land Degradation & Development **31:** 1380-1391, doi: 10.1002/ldr.3556

Hudson RG, Lipka J (1917) Padé 2/1 (Hudson-Lipka) elliptical perimeter approximation. *In*,

Kirschbaum MUF, McMillan AMS (2018) Warming and elevated CO2 have opposing influences on transpiration. Which is more important? Current Forestry Reports **4:** 51-71, doi: 10.1007/s40725-018-0073-8

Krejcia C, Beamon B (2015) Impacts of farmer coordination decisions on food supply chain structure Journal of Artificial Societies and Social Simulation **18** doi: 10.18564/jasss.2727

Lin Y (2015) LiDAR: An important tool for next-generation phenotyping technology of high potential for plant phenomics? Comput. Electron. Agric. **119:** 61-73, doi: 10.1016/j.compag.2015.10.011

Lucena LRRd, Leite MLdMV, Simões VJLP, Nóbrega C, Almeida MCR, Simplício JB (2021) Estimating the area and weight of cactus forage cladodes using linear dimensions Acta Scientarium Agronomy **43** doi: 10.4025/actasciagron.v43i1.45460

Neupane D, Mayer JA, Niechayev NA, Bishop CD, Cushman JC (2021) Five-year field trial of the biomass productivity and water input response of cactus pear (*Opuntia* spp.) as a bioenergy feedstock for arid lands. GCB Bioenergy **13:** 719-741, doi: 10.1111/gcbb.12805

Neupane D, Mayer JA, Niechayev NA, Bishop CD, Cushman JC (2023) Biomass production potential of 14 accessions of cactus pear (*Opuntia* spp.) as a food, feed, and biofuel crop for arid lands. *In*, Journal of Agronomy and Crop Science

Organisation for Economic Co-operation and Development (2022) What governments are doing. **Access 2022:** https://www.oecd.org/agriculture/understanding-the-global-food-system/what-governments-are-doing/

Pinheiro KM, Silva TGFd, Diniz WJdS, Carvalho HFdS, Moura MSBd (2015) Indirect methods for determining the area index of forage cactus cladodes. Pesquisa Agropecuária Tropical **45** doi: 10.1590/1983-40632015v4530617

R Development Core Team (2010) R: A language and environment for statistical computing. *In*. R Foundation for Statistical Computing

Reis CMG, Gazarini LC, Fonseca TF, Ribeiro MM (2018) Above-ground biomass estimation of *Opuntia ficus-indica* (L.) mill. for forage crop in a mediterranean environment by using non-destructive methods. Experimental Agriculture **54:** 227-242, doi: 10.1017/S0014479716000211

Reis CMG, Ribeiro MM, Gazarini LC (2014) Lecture: Caracterización y evaluación de las poblaciones portuguesas de chumbera *Opuntia ficus-indica* (L.) Mill. *In*, Repositório Universidade de Évoradoi: http://hdl.handle.net/10174/19753

Science PSECo (2022) STAT 502 - 11.2 - Using leverages to help identify extreme x values. **Access 2022:** https://online.stat.psu.edu/stat501/lesson/11/11.2

Şen Z (2009) Global warming threat on water resources and environment: a review. Environmental Geology **57:** 321-329, doi: 10.1007/s00254-008-1569-5

Sykora; S (2005) Approximations of ellipse perimeters and of the complete elliptic integral E(x). Review of known formulae Extra Byte **1**doi: 10.3247/SL1Math05.004

Thapa S, Zhu F, Walia H, Yu H, Ge Y (2018) A Novel LiDAR-Based Instrument for High-Throughput, 3D Measurement of Morphological Traits in Maize and Sorghum. Sensors **18:** 1187, doi: 10.3390/s18041187

Tiznado-Hernández ME, Fortiz-Hernández J, Ojeda-Contreras ÁJ, Rodríguez-Félix A (2010) Use of the elliptical mathematical formula to estimate the surface area of cladodes in four varieties of *Opuntia ficus-indica*. *In*,

USDA Natural Resources Conservation Service (2022) Web Soil Survey. **Access 2022:** https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx

Vaughan JE (2015) Investigating the genetic basis of natural leaf shape variation in *Arabidopsis thaliana*. *In*,

Appendices

A1. Table of recorded area values by cactus pad, as measured in ImageJ. Area units are square centimeters.

CompleteID	Area
PARL 580-1-1	417.132
PARL 580-1-2	643.954
PARL 580-1-3	497.155
PARL 580-1-4	580.734
PARL 580-1-5	581.399
PARL 580-1-6	377.744
PARL 580-1-7	415.3
PARL 580-1-8	600.257
PARL 854-1-1	412.126
PARL 854-1-2	305.532
PARL 854-1-3	376.755
PARL 854-1-4	600.126
PARL 854-1-5	395.247
PARL 854-1-6	369.877
PARL 854-1-7	343.71
PARL 854-1-8	310.113
PARL 584-1-1	662.101
PARL 584-1-2	1082.723
PARL 584-1-3	411.257
PARL 584-1-4	441.908
PARL 584-1-5	637.628
PARL 584-1-6	790.214
PARL 584-1-7	661.703
PARL 584-1-8	794.664
PARL 585-1-1	489.187
PARL 585-1-2	532.337
PARL 585-1-3	781.788
PARL 585-1-4	376.584
PARL 585-1-5	348.011
PARL 585-1-6	709.003
PARL 585-1-7	566.684
PARL 585-1-8	501.842
PARL 839-1-1	1077.263
PARL 839-1-2	764.287
PARL 839-1-3	810.49
PARL 839-1-4	656.562
PARL 839-1-5	769.322
PARL 839-1-6	647.915

PARL 839-1-7	754.456
PARL 839-1-8	721.762
PARL 325-1-1	460.352
PARL 325-1-2	821.59
PARL 325-1-3	589.758
PARL 325-1-4	691.811
PARL 325-1-5	740.908
PARL 325-1-6	879.41
PARL 325-1-7	665.612
PARL 325-1-8	794.164
PARL 572-1-1	513.687
PARL 572-1-2	798.06
PARL 572-1-3	462.863
PARL 572-1-4	622.495
PARL 572-1-5	484.885
PARL 572-1-6	363.522
PARL 572-1-7	312.859
PARL 572-1-8	414.154
PARL 390-1-1	595.133
PARL 390-1-2	573.089
PARL 390-1-3	549.63
PARL 390-1-4	679.157
PARL 390-1-5	362.294
PARL 390-1-6	939.224
PARL 390-1-7	590.339
PARL 390-1-8	712.298
PARL 319-1-1	1016.323
PARL 319-1-2	558.622
PARL 319-1-3	511.394
PARL 319-1-4	444.983
PARL 319-1-5	768
PARL 319-1-6	692.703
PARL 319-1-7	713.856
PARL 319-1-8	643.573
PARL 326-1-1	507.246
PARL 326-1-2	286.875
PARL 326-1-3	315.331
PARL 326-1-4	427.512
PARL 326-1-5	513.085
PARL 326-1-6	600.335
PARL 326-1-7	392.06
PARL 326-1-8	473.683
PARL 845-1-1	643.653

PARL 845-1-2	386.661
PARL 845-1-3	499.65
PARL 845-1-4	474.802
PARL 845-1-5	334.149
PARL 845-1-6	411.551
PARL 845-1-7	572.141
PARL 845-1-8	603.515
PARL 582-1-1	509.709
PARL 582-1-2	562.682
PARL 582-1-3	639.078
PARL 582-1-4	504.385
PARL 582-1-5	469.361
PARL 582-1-6	446.743
PARL 582-1-7	452.916
PARL 582-1-8	508.704
PARL 242-1-1	348.086
PARL 242-1-2	220.588
PARL 242-1-3	255.127
PARL 242-1-4	331.623
PARL 242-1-5	278.28
PARL 242-1-6	176.906
PARL 242-1-7	197.491
PARL 242-1-8	262.054
PARL 246-1-1	229.317
PARL 246-1-2	371.976
PARL 246-1-3	350.19
PARL 246-1-4	337.208
PARL 246-1-5	490.069
PARL 246-1-6	313.353
PARL 246-1-7	285.736
PARL 246-1-8	538.881
PARL 580-2-1	443.883
PARL 580-2-2	448.054
PARL 580-2-3	451.65
PARL 580-2-4	440.46
PARL 580-2-5	524.433
PARL 580-2-6	545.727
PARL 580-2-7	451.851
PARL 580-2-8	495.297
PARL 854-2-1	318.825
PARL 854-2-2	202.097
PARL 854-2-3	331.871
PARL 854-2-4	283.58

PARL 854-2-5	366.642
PARL 854-2-6	314.477
PARL 854-2-7	258.384
PARL 854-2-8	298.197
PARL 584-2-1	629.527
PARL 584-2-2	430.925
PARL 584-2-3	576.008
PARL 584-2-4	553.295
PARL 584-2-5	301.912
PARL 584-2-6	400.908
PARL 584-2-7	387.128
PARL 584-2-8	398.771
PARL 585-2-1	415.206
PARL 585-2-2	499.762
PARL 585-2-3	367.564
PARL 585-2-4	476.465
PARL 585-2-5	356.023
PARL 585-2-6	440.719
PARL 585-2-7	272.125
PARL 585-2-8	393.38
PARL 839-2-8	792.111
PARL 839-2-7	382.341
PARL 839-2-6	607.124
PARL 839-2-5	850.444
PARL 839-2-4	628.393
PARL 839-2-3	323.411
PARL 839-2-2	593.937
PARL 839-2-1	401.356
PARL 325-2-8	873.882
PARL 325-2-7	610.439
PARL 325-2-6	822.722
PARL 325-2-5	658.074
PARL 325-2-4	427.856
PARL 325-2-3	735.817
PARL 325-2-2	832.178
PARL 325-2-1	917.977
PARL 572-2-8	540.559
PARL 572-2-7	533.472
PARL 572-2-6	395.54
PARL 572-2-5	458.627
PARL 572-2-4	443.798
PARL 572-2-3	503.087
PARL 572-2-2	488.408

PARL 572-2-1	538.283
PARL 390-2-8	712.049
PARL 390-2-7	525.266
PARL 390-2-6	360.734
PARL 390-2-5	944.334
PARL 390-2-4	652.724
PARL 390-2-3	652.624
PARL 390-2-2	642.052
PARL 390-2-1	714.156
PARL 319-2-8	592.967
PARL 319-2-7	701.061
PARL 319-2-6	582.961
PARL 319-2-5	407.113
PARL 319-2-4	604.193
PARL 319-2-3	614.857
PARL 319-2-2	536.166
PARL 319-2-1	555.725
PARL 326-2-8	421.613
PARL 326-2-7	299.438
PARL 326-2-6	349.345
PARL 326-2-5	552.097
PARL 326-2-4	322.372
PARL 326-2-3	474.086
PARL 326-2-2	362.016
PARL 326-2-1	478.74
PARL 845-2-8	612.349
PARL 845-2-7	894.717
PARL 845-2-6	321.521
PARL 845-2-5	355.805
PARL 845-2-4	536.732
PARL 845-2-3	782.938
PARL 845-2-2	652.543
PARL 845-2-1	299.143
PARL 582-2-8	600.101
PARL 582-2-7	592.73
PARL 582-2-6	529.949
PARL 582-2-5	469.552
PARL 582-2-4	520.518
PARL 582-2-3	513.951
PARL 582-2-2	517.594
PARL 582-2-1	486.989
PARL 242-2-8	470.548
PARL 242-2-7	458.342

PARL 242-2-6	208.288
PARL 242-2-5	248.481
PARL 242-2-4	244.015
PARL 242-2-3	257.513
PARL 242-2-2	268.086
PARL 242-2-1	348.028
PARL 246-2-1	279.218
PARL 246-2-2	332.406
PARL 246-2-3	280.901
PARL 246-2-4	408.747
PARL 246-2-5	209.728
PARL 246-2-6	551.318
PARL 246-2-7	318.554
PARL 246-2-8	341.545
PARL 580-3-8	366.637
PARL 580-3-7	339.302
PARL 580-3-6	623.343
PARL 580-3-5	361.437
PARL 580-3-4	467.97
PARL 580-3-3	511.1
PARL 580-3-2	332.004
PARL 580-3-1	553.118
PARL 854-3-8	320.005
PARL 854-3-7	227.117
PARL 854-3-6	292.369
PARL 854-3-5	262.108
PARL 854-3-4	347.241
PARL 854-3-3	296.809
PARL 854-3-2	286.012
PARL 854-3-1	381.941
PARL 584-3-8	368.712
PARL 584-3-7	394.973
PARL 584-3-6	626.891
PARL 584-3-5	505.871
PARL 584-3-4	293.941
PARL 584-3-3	468.15
PARL 584-3-2	713.45
PARL 584-3-1	929.859
PARL 585-3-1	576.333
PARL 585-3-2	506.076
PARL 585-3-3	477.378
PARL 585-3-4	497.15
PARL 585-3-5	399.434

PARL 585-3-6	485.501
PARL 585-3-7	431.553
PARL 585-3-8	510.698
PARL 839-3-1	541.234
PARL 839-3-2	705.48
PARL 839-3-3	704.171
PARL 839-3-4	591.546
PARL 839-3-5	689.956
PARL 839-3-6	702.965
PARL 839-3-7	600
PARL 839-3-8	628.918
PARL 325-3-1	873.507
PARL 325-3-2	1003.913
PARL 325-3-3	876.808
PARL 325-3-4	921.74
PARL 325-3-5	703.941
PARL 325-3-6	860.859
PARL 325-3-7	586.023
PARL 325-3-8	804.08
PARL 572-3-1	567.712
PARL 572-3-2	493.501
PARL 572-3-3	351.851
PARL 572-3-4	511.833
PARL 572-3-5	634.076
PARL 572-3-6	555.464
PARL 572-3-7	495.398
PARL 572-3-8	553.382
PARL 390-3-1	621.012
PARL 390-3-2	814.715
PARL 390-3-3	653.365
PARL 390-3-4	494.991
PARL 390-3-5	778.468
PARL 390-3-6	844.605
PARL 390-3-7	653.224
PARL 390-3-8	701.562
PARL 319-3-1	698.025
PARL 319-3-2	712.356
PARL 319-3-3	530.516
PARL 319-3-4	610.738
PARL 319-3-5	659.612
PARL 319-3-6	528.504
PARL 319-3-7	664.208
PARL 319-3-8	659.167

PARL 326-3-1	370.314
PARL 326-3-2	528.902
PARL 326-3-3	498.697
PARL 326-3-4	323.897
PARL 326-3-5	387.221
PARL 326-3-6	476.175
PARL 326-3-7	501.636
PARL 326-3-8	384.254
PARL 845-3-1	615.582
PARL 845-3-2	503.196
PARL 845-3-3	437.795
PARL 845-3-4	677.038
PARL 845-3-5	438.499
PARL 845-3-6	582.186
PARL 845-3-7	628.733
PARL 845-3-8	807.277
PARL 582-3-1	515.272
PARL 582-3-2	571.011
PARL 582-3-3	419.424
PARL 582-3-4	423.618
PARL 582-3-5	588.555
PARL 582-3-6	526.965
PARL 582-3-7	508.929
PARL 582-3-8	479.747
PARL 242-3-1	238.771
PARL 242-3-2	489.488
PARL 242-3-3	406.703
PARL 242-3-4	291.273
PARL 242-3-5	306.453
PARL 242-3-6	388.259
PARL 242-3-7	434.543
PARL 242-3-8	260.032
PARL 246-3-1	346.133
PARL 246-3-2	246.713
PARL 246-3-3	307.079
PARL 246-3-4	414.577
PARL 246-3-5	337.507
PARL 246-3-6	439.32
PARL 246-3-7	334.573
PARL 246-3-8	303.164
PARL 580-4-1	403.214
PARL 580-4-2	368.488
PARL 580-4-3	543.162

PARL 580-4-4	473.147
PARL 580-4-5	446.622
PARL 580-4-6	449.65
PARL 580-4-7	424.643
PARL 580-4-8	520.814
PARL 854-4-1	229.442
PARL 854-4-2	232.04
PARL 854-4-3	219.486
PARL 854-4-4	260.345
PARL 854-4-5	364.738
PARL 854-4-6	342.698
PARL 854-4-7	240.333
PARL 854-4-8	233.917
PARL 584-4-1	490.124
PARL 584-4-2	479.094
PARL 584-4-3	301.5
PARL 584-4-4	622.138
PARL 584-4-5	598.181
PARL 584-4-6	741.041
PARL 584-4-7	404.368
PARL 584-4-8	485.001
PARL 585-4-1	437.01
PARL 585-4-2	439.991
PARL 585-4-3	242.276
PARL 585-4-4	321.793
PARL 585-4-5	449.483
PARL 585-4-6	495.716
PARL 585-4-7	406.453
PARL 585-4-8	423.686
PARL 839-4-1	735.663
PARL 839-4-2	578.965
PARL 839-4-3	509.909
PARL 839-4-4	767.38
PARL 839-4-5	410.874
PARL 839-4-6	594.274
PARL 839-4-7	NA
PARL 839-4-8	414.758
PARL 325-4-1	613.783
PARL 325-4-2	690.986
PARL 325-4-3	723.111
PARL 325-4-4	668.276
PARL 325-4-5	641.542
PARL 325-4-6	591.612

PARL 325-4-7	666.627
PARL 325-4-8	1073.268
PARL 572-4-1	456.518
PARL 572-4-2	467.439
PARL 572-4-3	541.929
PARL 572-4-4	559.146
PARL 572-4-5	580.476
PARL 572-4-6	457.699
PARL 572-4-7	560.824
PARL 572-4-8	698.503
PARL 390-4-1	867.278
PARL 390-4-2	502.354
PARL 390-4-3	582.549
PARL 390-4-4	385.908
PARL 390-4-5	649.998
PARL 390-4-6	517.946
PARL 390-4-7	876.922
PARL 390-4-8	604.245
PARL 319-4-8	689.465
PARL 319-4-7	609.833
PARL 319-4-6	468.784
PARL 319-4-5	297.914
PARL 319-4-4	393.432
PARL 319-4-3	587.955
PARL 319-4-2	575.986
PARL 319-4-1	606.361
PARL 326-4-8	382.306
PARL 326-4-7	555.781
PARL 326-4-6	428.388
PARL 326-4-5	473.002
PARL 326-4-4	453.161
PARL 326-4-3	426.017
PARL 326-4-2	329.284
PARL 326-4-1	527.801
PARL 845-4-8	356.23
PARL 845-4-7	448.892
PARL 845-4-6	382.758
PARL 845-4-5	288.803
PARL 845-4-4	291.298
PARL 845-4-3	497.725
PARL 845-4-2	437.214
PARL 845-4-1	313.087
PARL 582-4-8	273.241

PARL 582-4-7	667.402
PARL 582-4-6	546.684
PARL 582-4-5	520.102
PARL 582-4-4	513.447
PARL 582-4-3	519.672
PARL 582-4-2	620.334
PARL 582-4-1	416.069
PARL 242-4-8	543.248
PARL 242-4-7	301.354
PARL 242-4-6	236.34
PARL 242-4-5	201.525
PARL 242-4-4	214.616
PARL 242-4-3	426.091
PARL 242-4-2	418.78
PARL 242-4-1	281.938
PARL 246-4-1	431.254
PARL 246-4-2	264.572
PARL 246-4-3	250.572
PARL 246-4-4	243.529
PARL 246-4-5	324.704
PARL 246-4-6	305.962
PARL 246-4-7	303.799
PARL 246-4-8	332.518

A2. Complete table of unaugmented input data. IDs, accession information, and species identification are included with the initial five measures of fresh weight, width, height, diameter, and thickness.

CompleteID	Fresh_weight	Width	Height	Diameter	Thickness	Accession	Species
PARL 242-1-1	515.2	14.4	33.1	75.6	17.3	242	*O. cochenillifera*
PARL 242-1-2	249.6	10	28.5	64.2	14	242	*O. cochenillifera*
PARL 242-1-3	291.3	8.7	33	73.3	16.3	242	*O. cochenillifera*
PARL 242-1-4	422.1	13.4	33.1	73.9	16.3	242	*O. cochenillifera*
PARL 242-1-5	366.9	10.5	34	74.5	17.1	242	*O. cochenillifera*
PARL 242-1-6	202.3	7.6	30.5	66.2	15.2	242	*O. cochenillifera*
PARL 242-1-7	244.6	7.3	21	69.2	17.4	242	*O. cochenillifera*
PARL 242-1-8	270.4	9.4	34.6	75.4	14.7	242	*O. cochenillifera*
PARL 242-2-1	446.6	14.2	33.4	76.6	16.1	242	*O. cochenillifera*
PARL 242-2-2	291.5	9.5	37.7	82.7	14.1	242	*O. cochenillifera*
PARL 242-2-3	363.1	10.7	30.5	70.5	17.3	242	*O. cochenillifera*
PARL 242-2-4	271.5	11.3	29.1	66.5	14.7	242	*O. cochenillifera*
PARL 242-2-5	327	9	35.7	77.4	17.1	242	*O. cochenillifera*
PARL 242-2-6	192.7	9.2	29.3	65.1	12.6	242	*O. cochenillifera*
PARL 242-2-7	901.8	15	41.9	94.6	23.3	242	*O. cochenillifera*
PARL 242-2-8	649.3	13.6	44.3	98.4	14.2	242	*O. cochenillifera*
PARL 242-3-1	279.1	11.4	29.6	66.1	13.8	242	*O. cochenillifera*
PARL 242-3-2	744.4	13.9	48.3	105.3	17.5	242	*O. cochenillifera*
PARL 242-3-3	627.2	16.3	33.6	80.5	15.6	242	*O. cochenillifera*
PARL 242-3-4	348.9	12.4	30.9	72	11.1	242	*O. cochenillifera*
PARL 242-3-5	374.7	12.8	32.9	74.6	13.3	242	*O. cochenillifera*
PARL 242-3-6	486.4	12.3	41.9	91.8	15.8	242	*O. cochenillifera*
PARL 242-3-7	581.4	13.1	44.7	98.3	13.7	242	*O. cochenillifera*
PARL 242-3-8	335.2	10.4	33.7	74.9	14	242	*O. cochenillifera*
PARL 242-4-1	352	11.6	31.7	72.6	13.1	242	*O. cochenillifera*
PARL 242-4-2	792.7	14.9	39.3	89.7	19	242	*O. cochenillifera*
PARL 242-4-3	558.9	13.3	41.2	94.1	12.5	242	*O. cochenillifera*
PARL 242-4-4	268	9.1	33.2	72.7	14.2	242	*O. cochenillifera*
PARL 242-4-5	235.9	8.4	30.6	67.7	15.4	242	*O. cochenillifera*
PARL 242-4-6	270.8	9.5	32.4	71.4	12.4	242	*O. cochenillifera*
PARL 242-4-7	352	10.1	38.6	84.6	13.7	242	*O. cochenillifera*
PARL 242-4-8	1149	18.2	41.3	96.1	22.1	242	*O. cochenillifera*
PARL 246-1-1	301	13.4	21.6	56.2	14.8	246	*O. ficus-indica*
PARL 246-1-2	595.4	14.5	34.1	78.3	20.3	246	*O. ficus-indica*
PARL 246-1-3	517	16.4	25.9	67.2	15.7	246	*O. ficus-indica*
PARL 246-1-4	475.9	16.3	28	70.1	15.9	246	*O. ficus-indica*
PARL 246-1-5	769.4	19.9	33.5	82.2	16.3	246	*O. ficus-indica*
PARL 246-1-6	361.1	12.6	31.1	71.4	12.3	246	*O. ficus-indica*
PARL 246-1-7	282.5	13.9	26.2	64.4	12.4	246	*O. ficus-indica*

PARL 246-1-8	747.7	20	31.6	82.5	14.6	246	*O. ficus-indica*
PARL 246-2-1	346	20	27.4	65	13.6	246	*O. ficus-indica*
PARL 246-2-2	426.5	15.6	30.6	73.5	15.7	246	*O. ficus-indica*
PARL 246-2-3	390.5	13.1	27.6	66.9	16.2	246	*O. ficus-indica*
PARL 246-2-4	609.9	18.1	29.3	75.5	15.2	246	*O. ficus-indica*
PARL 246-2-5	295.5	10.8	25.4	60.2	19.5	246	*O. ficus-indica*
PARL 246-2-6	807.8	20.2	37.6	92.1	13.7	246	*O. ficus-indica*
PARL 246-2-7	366	12.2	34.2	77.8	15.4	246	*O. ficus-indica*
PARL 246-2-8	416.4	13.4	36.4	79.3	13	246	*O. ficus-indica*
PARL 246-3-1	447	16.6	30.4	73.4	14.4	246	*O. ficus-indica*
PARL 246-3-2	292.5	14.3	23.6	56.4	12.7	246	*O. ficus-indica*
PARL 246-3-3	388.3	13.5	32.6	74.7	13	246	*O. ficus-indica*
PARL 246-3-4	503.7	16.4	35.2	83	13.2	246	*O. ficus-indica*
PARL 246-3-5	417.9	16.6	28.1	70.1	13.2	246	*O. ficus-indica*
PARL 246-3-6	604.6	18.4	33.5	81.9	15	246	*O. ficus-indica*
PARL 246-3-7	459.5	13.7	32.7	78.3	14.2	246	*O. ficus-indica*
PARL 246-3-8	452.7	15.4	28.6	68.3	16.7	246	*O. ficus-indica*
PARL 246-4-1	790.7	17.6	36	85	15.8	246	*O. ficus-indica*
PARL 246-4-2	434.4	15.6	25.8	64.8	19.1	246	*O. ficus-indica*
PARL 246-4-3	300.8	13.2	27.4	64.1	13.9	246	*O. ficus-indica*
PARL 246-4-4	341.4	11.6	30.1	67.3	14	246	*O. ficus-indica*
PARL 246-4-5	530.7	15.2	29.5	72.1	19.4	246	*O. ficus-indica*
PARL 246-4-6	336.1	14.1	30.5	71.6	10.7	246	*O. ficus-indica*
PARL 246-4-7	479.3	12.8	34.2	76.7	16.8	246	*O. ficus-indica*
PARL 246-4-8	534.6	16.2	28.1	70.5	15.7	246	*O. ficus-indica*
PARL 319-1-1	1591.2	19.2	52.3	12.3	20.3	319	*O. cochenillifera*
PARL 319-1-2	881.1	15.1	37.6	85.8	23.7	319	*O. cochenillifera*
PARL 319-1-3	673.2	13.3	39.1	89	19.2	319	*O. cochenillifera*
PARL 319-1-4	627.4	12.2	35.2	81.9	21	319	*O. cochenillifera*
PARL 319-1-5	1204.7	14.4	17.5	102.8	25.5	319	*O. cochenillifera*
PARL 319-1-6	920.5	16.5	42	96.6	20	319	*O. cochenillifera*
PARL 319-1-7	1227.9	15.5	43.8	98.6	27.6	319	*O. cochenillifera*
PARL 319-1-8	830.5	14.6	44.6	99.8	21	319	*O. cochenillifera*
PARL 319-2-1	1050.7	17.3	45.2	100.5	23.5	319	*O. cochenillifera*
PARL 319-2-2	971.5	16.4	43.2	98.9	22.2	319	*O. cochenillifera*
PARL 319-2-3	1210.2	19.3	44.8	103.2	24.7	319	*O. cochenillifera*
PARL 319-2-4	1230.7	15.6	52.8	115.9	23.1	319	*O. cochenillifera*
PARL 319-2-5	1002.5	12.4	47.4	102	32.1	319	*O. cochenillifera*
PARL 319-2-6	1134.4	16.4	48.8	107.4	26.1	319	*O. cochenillifera*
PARL 319-2-7	1661.8	17.8	52.6	114.8	26.7	319	*O. cochenillifera*
PARL 319-2-8	1285.2	16.2	47.3	107.4	29.1	319	*O. cochenillifera*
PARL 319-3-1	1420.3	17.2	56.1	123.4	20.9	319	*O. cochenillifera*
PARL 319-3-2	1259.9	19.9	48.5	111.3	23.6	319	*O. cochenillifera*

PARL 319-3-3	942.8	13.8	51.7	111.8	21.3	319	*O. cochenillifera*
PARL 319-3-4	1350	16.1	51.3	115.3	27.9	319	*O. cochenillifera*
PARL 319-3-5	1221.7	17.7	43.5	105.9	18.7	319	*O. cochenillifera*
PARL 319-3-6	961.6	15.3	46.1	104	25.1	319	*O. cochenillifera*
PARL 319-3-7	1253.9	17.1	46.6	105.5	22.1	319	*O. cochenillifera*
PARL 319-3-8	1310.2	18.7	48.6	108.2	24.7	319	*O. cochenillifera*
PARL 319-4-1	1220.6	15.4	54.1	117.6	29.2	319	*O. cochenillifera*
PARL 319-4-2	1009.1	17.1	46.4	105.5	20.6	319	*O. cochenillifera*
PARL 319-4-3	1064.7	16.4	47.2	106	20.1	319	*O. cochenillifera*
PARL 319-4-4	729.7	14.1	38.6	78	23.6	319	*O. cochenillifera*
PARL 319-4-5	495.2	11	35.1	80	21.1	319	*O. cochenillifera*
PARL 319-4-6	974.7	12.2	50.2	106.6	21.3	319	*O. cochenillifera*
PARL 319-4-7	1139.2	18.6	41.4	101.1	23.1	319	*O. cochenillifera*
PARL 319-4-8	1410.1	16.2	61.3	123.6	20.7	319	*O. cochenillifera*
PARL 325-1-1	513.6	17.8	28.4	66.4	16	325	*O. crassa*
PARL 325-1-2	1813.7	23.9	48.7	113.3	25	325	*O. crassa*
PARL 325-1-3	1211.5	18.2	45.2	100.3	18	325	*O. crassa*
PARL 325-1-4	1390	24.3	35.6	92.8	22	325	*O. crassa*
PARL 325-1-5	1284.8	22.1	42.7	103.6	20	325	*O. crassa*
PARL 325-1-6	1781.3	26.4	44.9	108.6	29	325	*O. crassa*
PARL 325-1-7	1251.5	23.2	37.1	92.1	22	325	*O. crassa*
PARL 325-1-8	1670.5	21.2	46.8	105.8	28	325	*O. crassa*
PARL 325-2-1	1926.8	29.2	43.6	113.2	19.5	325	*O. crassa*
PARL 325-2-2	1667.6	27.4	42	104.4	19.8	325	*O. crassa*
PARL 325-2-3	1693.9	23.2	40.5	101.5	27.4	325	*O. crassa*
PARL 325-2-4	835.2	17.3	35.1	82.4	22.2	325	*O. crassa*
PARL 325-2-5	1165.9	20.5	43.2	100.5	22.9	325	*O. crassa*
PARL 325-2-6	1611.4	26.6	44.2	108.8	21.2	325	*O. crassa*
PARL 325-2-7	1239.9	22.8	36.4	91.3	22.2	325	*O. crassa*
PARL 325-2-8	2069.6	25	43.5	108	27.3	325	*O. crassa*
PARL 325-3-1	1826.1	26.1	48.4	117.3	19.6	325	*O. crassa*
PARL 325-3-2	2529.6	28.5	48.8	118.7	27.7	325	*O. crassa*
PARL 325-3-3	1957.9	25.4	44.6	110.7	27	325	*O. crassa*
PARL 325-3-4	2055.9	26.3	43.6	112.4	20.5	325	*O. crassa*
PARL 325-3-5	1440.5	19.5	46.4	109.6	25	325	*O. crassa*
PARL 325-3-6	1966.4	25.4	40	106.4	26.4	325	*O. crassa*
PARL 325-3-7	1256	18.3	39.1	94.6	25.8	325	*O. crassa*
PARL 325-3-8	1758.2	25.3	42.8	105.6	26.3	325	*O. crassa*
PARL 325-4-1	1272.3	20.8	40.7	96.3	22.4	325	*O. crassa*
PARL 325-4-2	1678.5	22.3	40.6	99.4	39.8	325	O. crassa
PARL 325-4-3	1226.1	23.8	39.7	101.3	18.3	325	O. crassa
PARL 325-4-4	1250.7	23.3	32.1	91.6	18.9	325	O. crassa
PARL 325-4-5	1244.2	23.9	38.8	94.4	23.9	325	O. crassa

PARL 325-4-6	1191.4	19.7	42.9	98.9	23.1	325	O. crassa
PARL 325-4-7	1490.1	20.4	46.8	106	23.4	325	O. crassa
PARL 325-4-8	2437.8	29.8	48.1	122.8	25.6	325	O. crassa
PARL 326-1-1	1298.5	19.4	33.6	83.1	32.7	326	*O. ficus-indica*
PARL 326-1-2	455.9	14	25.2	62.7	20.4	326	*O. ficus-indica*
PARL 326-1-3	542.4	14.6	27.9	68.1	21.2	326	*O. ficus-indica*
PARL 326-1-4	908.7	17.1	32.3	78.9	26.6	326	*O. ficus-indica*
PARL 326-1-5	1163	18.4	34.4	83.8	26.3	326	*O. ficus-indica*
PARL 326-1-6	1445.3	19.3	38.9	94.7	29.7	326	*O. ficus-indica*
PARL 326-1-7	855.3	17.4	28.1	72.6	25.5	326	*O. ficus-indica*
PARL 326-1-8	1041.3	17.2	35.5	82.6	25.9	326	*O. ficus-indica*
PARL 326-2-1	892.1	18.2	34.8	82.8	20.3	326	*O. ficus-indica*
PARL 326-2-2	744.8	16.4	28.7	69.1	26.8	326	*O. ficus-indica*
PARL 326-2-3	893.1	17.5	30.5	80.4	21.6	326	*O. ficus-indica*
PARL 326-2-4	592	15.4	28.2	66.1	21.3	326	*O. ficus-indica*
PARL 326-2-5	1177.3	19.6	33	86.2	24.6	326	*O. ficus-indica*
PARL 326-2-6	783.5	16.1	29.7	70.8	31.9	326	*O. ficus-indica*
PARL 326-2-7	607.4	14.3	24.5	63.6	26	326	*O. ficus-indica*
PARL 326-2-8	809.7	18.1	31.6	76.9	24.4	326	*O. ficus-indica*
PARL 326-3-1	695.9	15.5	31.4	73.9	21.6	326	*O. ficus-indica*
PARL 326-3-2	1097	18.9	36.6	88.2	22.7	326	*O. ficus-indica*
PARL 326-3-3	1094.5	19.9	32.1	81.1	29.1	326	*O. ficus-indica*
PARL 326-3-4	628.1	16.5	25.1	65.8	24.1	326	*O. ficus-indica*
PARL 326-3-5	785	16.3	29.9	74.2	23.4	326	*O. ficus-indica*
PARL 326-3-6	913.3	18.7	30.9	80.3	24.8	326	*O. ficus-indica*
PARL 326-3-7	953	19.4	34.8	82.7	21.4	326	*O. ficus-indica*
PARL 326-3-8	831.7	14.9	31.3	75.7	24	326	*O. ficus-indica*
PARL 326-4-1	1224.1	19.2	35.9	88.6	28	326	*O. ficus-indica*
PARL 326-4-2	788	14.1	32.9	76.1	13	326	*O. ficus-indica*
PARL 326-4-3	906.6	18	32.9	79.1	27	326	*O. ficus-indica*
PARL 326-4-4	1073.4	18.5	33.7	80.6	30.6	326	*O. ficus-indica*
PARL 326-4-5	1012.2	19.7	29.6	78.1	25.2	326	*O. ficus-indica*
PARL 326-4-6	924.8	17.6	31.4	77.3	27.1	326	*O. ficus-indica*
PARL 326-4-7	1233.6	19.5	36.7	90.5	24.1	326	*O. ficus-indica*
PARL 326-4-8	783.5	17.2	31.7	74.3	25.4	326	*O. ficus-indica*
PARL 390-1-1	1016	21.9	34.3	89.9	18.5	390	*O. robusta*
PARL 390-1-2	1196.1	25	28.5	82.5	25.1	390	*O. robusta*
PARL 390-1-3	1043	22.1	29.1	80.8	26.1	390	*O. robusta*
PARL 390-1-4	1162.8	25.1	33.4	92.3	20.6	390	*O. robusta*
PARL 390-1-5	421.5	17.5	23.7	66.6	13.9	390	*O. robusta*
PARL 390-1-6	1512	29.5	38.8	107.2	21.5	390	*O. robusta*
PARL 390-1-7	850.3	23.5	30.1	84.5	15.7	390	*O. robusta*
PARL 390-1-8	1615.4	24.2	35.6	95.4	33.4	390	*O. robusta*

PARL 390-2-1	1539.5	27.8	34.8	95.7	23.8	390	*O. robusta*
PARL 390-2-2	1337.4	23.7	36	94.9	22.7	390	*O. robusta*
PARL 390-2-3	1309.4	25.5	36	100.4	24.2	390	*O. robusta*
PARL 390-2-4	1869.8	24	34.3	93.4	32.1	390	*O. robusta*
PARL 390-2-5	2263.4	33.4	38.7	109.9	25.9	390	*O. robusta*
PARL 390-2-6	1506.3	26.5	33	99.5	21.3	390	*O. robusta*
PARL 390-2-7	1160.9	21.6	31.1	82.5	27	390	*O. robusta*
PARL 390-2-8	1619.8	25.8	34.6	94.6	28.2	390	*O. robusta*
PARL 390-3-1	1214	24.6	35.8	92.9	20.1	390	*O. robusta*
PARL 390-3-2	1859.8	28.7	36.7	102.8	25.2	390	*O. robusta*
PARL 390-3-3	1407.3	22.8	34.5	93.9	24	390	*O. robusta*
PARL 390-3-4	934.5	22.5	27.7	78.1	19.9	390	*O. robusta*
PARL 390-3-5	1748.1	26.8	36.5	98.3	26.7	390	*O. robusta*
PARL 390-3-6	2064.4	30.7	37.2	105.7	27	390	*O. robusta*
PARL 390-3-7	1496.6	21.6	41.1	98.9	26.5	390	*O. robusta*
PARL 390-3-8	1323.7	24.4	36.5	96	18.9	390	*O. robusta*
PARL 390-4-1	2169.4	29.2	36.7	104.9	26.3	390	*O. robusta*
PARL 390-4-2	772.7	21.3	29.3	81.6	18.3	390	*O. robusta*
PARL 390-4-3	1380.5	26.4	27.1	86.2	27	390	*O. robusta*
PARL 390-4-4	657.8	20.3	22.4	70.6	20.7	390	*O. robusta*
PARL 390-4-5	1633.9	25.7	34.1	92.6	29.9	390	*O. robusta*
PARL 390-4-6	1019.6	24.8	28.2	58.6	24.6	390	*O. robusta*
PARL 390-4-7	1590.8	32.7	34.7	105.1	21	390	*O. robusta*
PARL 390-4-8	1194.9	23.3	33.4	90.7	23.9	390	*O. robusta*
PARL 572-1-1	854	19.2	35.5	84.4	17	572	*O. ficus-indica*
PARL 572-1-2	1193.4	21.7	45	105.1	17	572	*O. ficus-indica*
PARL 572-1-3	705.8	17.7	30.5	78	19	572	*O. ficus-indica*
PARL 572-1-4	1051.8	20.8	41.3	96.3	21	572	*O. ficus-indica*
PARL 572-1-5	868.2	19.4	31.8	79.7	23.4	572	*O. ficus-indica*
PARL 572-1-6	639.5	14.4	30	73.5	22	572	*O. ficus-indica*
PARL 572-1-7	527.1	14	28	68	19.8	572	*O. ficus-indica*
PARL 572-1-8	548	15.4	30.3	72.6	13	572	*O. ficus-indica*
PARL 572-2-1	808.2	19.6	38.3	91.7	14	572	*O. ficus-indica*
PARL 572-2-2	797.1	18.8	36.7	83.7	20.7	572	*O. ficus-indica*
PARL 572-2-3	817.1	19.7	36	86.2	14	572	*O. ficus-indica*
PARL 572-2-4	590.1	18	31	79.2	17	572	*O. ficus-indica*
PARL 572-2-5	726.7	17.7	36.8	84.7	19.1	572	*O. ficus-indica*
PARL 572-2-6	693.6	19.2	38	81.8	22.2	572	*O. ficus-indica*
PARL 572-2-7	956.5	19.8	38.6	89.2	23.3	572	*O. ficus-indica*
PARL 572-2-8	763.9	18.3	36.8	90.4	16.5	572	*O. ficus-indica*
PARL 572-3-1	943.3	19.7	42.2	96.9	18.6	572	*O. ficus-indica*
PARL 572-3-2	687.4	18.6	37.1	86.8	14.7	572	*O. ficus-indica*
PARL 572-3-3	607.7	15.2	30.6	73.3	20.3	572	*O. ficus-indica*

PARL 572-3-4	636.5	20.4	35.9	88.2	13.4	572	*O. ficus-indica*
PARL 572-3-5	1170.1	20.3	41	98.6	19.8	572	*O. ficus-indica*
PARL 572-3-6	927.6	21.4	40.4	91.7	16.1	572	*O. ficus-indica*
PARL 572-3-7	909.2	18.5	35.1	85.5	22.1	572	*O. ficus-indica*
PARL 572-3-8	1004.1	22.2	32.9	86.4	20	572	*O. ficus-indica*
PARL 572-4-1	748.7	19.2	37.3	84.1	16.5	572	*O. ficus-indica*
PARL 572-4-2	779.3	17.8	33.7	83.6	20.8	572	*O. ficus-indica*
PARL 572-4-3	1026.1	21.3	40.5	90.2	20.6	572	*O. ficus-indica*
PARL 572-4-4	1299.2	18.4	40.2	96.1	25.1	572	*O. ficus-indica*
PARL 572-4-5	1038.3	20.2	40.8	95.5	18.7	572	*O. ficus-indica*
PARL 572-4-6	901.1	19.8	34.3	83.3	25.3	572	*O. ficus-indica*
PARL 572-4-7	1164.7	20.6	37.2	90.6	20.9	572	*O. ficus-indica*
PARL 572-4-8	1257	22.6	43.8	105	18.3	572	*O. ficus-indica*
PARL 580-1-1	707.8	15	41.5	92.3	19.19	580	*O. ficus-indica*
PARL 580-1-2	1151.7	17.4	47.3	180.9	17.58	580	*O. ficus-indica*
PARL 580-1-3	821.8	17.4	42	94.4	17.48	580	*O. ficus-indica*
PARL 580-1-4	1135.4	17.5	46	100.5	22.78	580	*O. ficus-indica*
PARL 580-1-5	1034.1	16.5	43	101	18.73	580	*O. ficus-indica*
PARL 580-1-6	677.8	11.9	46.1	96.9	24.3	580	*O. ficus-indica*
PARL 580-1-7	714	14.7	39	82.5	16.99	580	*O. ficus-indica*
PARL 580-1-8	962.2	15.4	48.1	107.8	21.84	580	*O. ficus-indica*
PARL 580-2-1	664.5	15.4	37.9	85.4	17.5	580	*O. ficus-indica*
PARL 580-2-2	635.6	13.9	42	93.5	16.8	580	*O. ficus-indica*
PARL 580-2-3	893.4	14.2	38.3	90.3	25	580	*O. ficus-indica*
PARL 580-2-4	727	13.8	38.5	89.4	20.4	580	*O. ficus-indica*
PARL 580-2-5	1019.1	14.3	43.5	101.2	27	580	*O. ficus-indica*
PARL 580-2-6	990.6	14.4	48	107.9	21	580	*O. ficus-indica*
PARL 580-2-7	848.8	14.3	43.9	97.6	18.9	580	*O. ficus-indica*
PARL 580-2-8	830.6	15.1	40	89.3	20.6	580	*O. ficus-indica*
PARL 580-3-1	893.4	15.2	46.7	103.1	17.6	580	*O. ficus-indica*
PARL 580-3-2	644.8	14.1	30.9	73.7	22.2	580	*O. ficus-indica*
PARL 580-3-3	1015.8	14.6	41.8	97	23.1	580	*O. ficus-indica*
PARL 580-3-4	759.1	16.6	38.2	87.2	20.6	580	*O. ficus-indica*
PARL 580-3-5	718.6	15.4	31.2	102.6	25.4	580	*O. ficus-indica*
PARL 580-3-6	1210	16.6	44.7	74.7	26.5	580	*O. ficus-indica*
PARL 580-3-7	552.9	11.6	38.6	83.8	20.6	580	*O. ficus-indica*
PARL 580-3-8	721.7	13	34.1	74.6	22.8	580	*O. ficus-indica*
PARL 580-4-1	848.7	15.1	36.4	81.6	25.7	580	*O. ficus-indica*
PARL 580-4-2	687.4	13.8	38.9	86.1	21.2	580	*O. ficus-indica*
PARL 580-4-3	925.9	17.7	42.2	96.8	20.8	580	*O. ficus-indica*
PARL 580-4-4	1106.6	16.6	37.7	87.8	27.5	580	*O. ficus-indica*
PARL 580-4-5	734.1	15.1	38.2	88.8	18	580	*O. ficus-indica*
PARL 580-4-6	979.4	16.4	36.6	91.1	23.1	580	*O. ficus-indica*

PARL 580-4-7	824.9	14.7	41.2	88.2	21.9	580	*O. ficus-indica*
PARL 580-4-8	1171.3	17.9	42.2	94.7	26.4	580	*O. ficus-indica*
PARL 582-1-1	973.9	15.6	40.2	93.2	22.5	582	*O. cochenillifera*
PARL 582-1-2	1055.7	13.6	49.5	108.8	23.7	582	*O. cochenillifera*
PARL 582-1-3	1140.2	18.1	43.9	102.7	22.7	582	*O. cochenillifera*
PARL 582-1-4	847.4	15.3	40	92.7	22.4	582	*O. cochenillifera*
PARL 582-1-5	680	14.1	39.3	87.8	19.8	582	*O. cochenillifera*
PARL 582-1-6	807.5	14.9	37.2	87.8	21.8	582	*O. cochenillifera*
PARL 582-1-7	811.6	13.9	41.5	93.6	21.8	582	*O. cochenillifera*
PARL 582-1-8	859.6	14.7	43.7	97.8	19.7	582	*O. cochenillifera*
PARL 582-2-1	842.5	15.7	43.1	96.4	22.2	582	*O. cochenillifera*
PARL 582-2-2	936.2	16.5	41.4	96.1	22.9	582	*O. cochenillifera*
PARL 582-2-3	986.9	16.7	40.6	94.6	22.6	582	*O. cochenillifera*
PARL 582-2-4	961.8	14.6	43	99.4	21.5	582	*O. cochenillifera*
PARL 582-2-5	744.6	14.3	43.1	95.8	21.8	582	*O. cochenillifera*
PARL 582-2-6	836.6	16.2	45.6	100.6	18.3	582	*O. cochenillifera*
PARL 582-2-7	1267.9	17.6	43	99.4	25.3	582	*O. cochenillifera*
PARL 582-2-8	897.2	16.8	42.4	99.3	20.3	582	*O. cochenillifera*
PARL 582-3-1	994.9	15.6	37.3	103.1	23.9	582	*O. cochenillifera*
PARL 582-3-2	1075.4	17.1	43.8	99.7	24.2	582	*O. cochenillifera*
PARL 582-3-3	852.9	13.6	45.7	100.2	23	582	*O. cochenillifera*
PARL 582-3-4	780.6	13.3	40.3	90.6	21.6	582	*O. cochenillifera*
PARL 582-3-5	1150.5	17.9	44.8	101.6	20.8	582	*O. cochenillifera*
PARL 582-3-6	1051.2	15.7	45.6	102.6	22.7	582	*O. cochenillifera*
PARL 582-3-7	1050.7	15.4	47.1	102.3	24.6	582	*O. cochenillifera*
PARL 582-3-8	863.1	14.2	43.6	97.4	20.3	582	*O. cochenillifera*
PARL 582-4-1	695.7	13.4	40.3	90.6	18.7	582	*O. cochenillifera*
PARL 582-4-2	1104.1	17.4	43.6	104.2	17.8	582	*O. cochenillifera*
PARL 582-4-3	970	16.4	44.2	97.8	23.6	582	*O. cochenillifera*
PARL 582-4-4	953.5	16.9	41.4	97.1	20	582	*O. cochenillifera*
PARL 582-4-5	1013.2	16.4	42.2	98.9	25.3	582	*O. cochenillifera*
PARL 582-4-6	1109.8	17.4	43.7	98.9	23.9	582	*O. cochenillifera*
PARL 582-4-7	1400.9	18.6	47.3	109.5	24.9	582	*O. cochenillifera*
PARL 582-4-8	477.1	11.7	32.2	67	19	582	*O. cochenillifera*
PARL 584-1-1	1465.2	22.3	36.2	228	27.18	584	*O. ficus-indica*
PARL 584-1-2	2323.9	28.5	44.5	115.5	23.5	584	*O. ficus-indica*
PARL 584-1-3	756.3	19.7	25.6	70.5	23.6	584	*O. ficus-indica*
PARL 584-1-4	839.8	16.5	31.3	78.2	24	584	*O. ficus-indica*
PARL 584-1-5	1348	21.7	37.2	92.1	26.17	584	*O. ficus-indica*
PARL 584-1-6	1800.2	20.5	43	105	26.5	584	*O. ficus-indica*
PARL 584-1-7	1462.6	20.4	42.3	98.5	25.77	584	*O. ficus-indica*
PARL 584-1-8	1609.2	20.5	44.5	107	25	584	*O. ficus-indica*
PARL 584-2-1	1107.9	21.7	36.4	91.6	18.7	584	*O. ficus-indica*

PARL 584-2-2	987.5	17.8	33.2	78.2	28.7	584	*O. ficus-indica*
PARL 584-2-3	1093.4	20.5	38	81.9	27	584	*O. ficus-indica*
PARL 584-2-4	1281.4	20.5	34.2	88.3	29.4	584	*O. ficus-indica*
PARL 584-2-5	542.2	14.2	26.2	65.9	19.9	584	*O. ficus-indica*
PARL 584-2-6	881.4	17	28.5	74.2	23.3	584	*O. ficus-indica*
PARL 584-2-7	840.4	14.6	31.6	76.7	31.1	584	*O. ficus-indica*
PARL 584-2-8	813	16.6	30.3	73.5	21.9	584	*O. ficus-indica*
PARL 584-3-1	1957.7	26.3	47.2	114.4	21.9	584	*O. ficus-indica*
PARL 584-3-2	1406.9	24.4	35.9	94.9	23.6	584	*O. ficus-indica*
PARL 584-3-3	859.2	18.8	31.2	78	23.7	584	*O. ficus-indica*
PARL 584-3-4	430.2	16.8	24.9	64.9	19.9	584	*O. ficus-indica*
PARL 584-3-5	927.7	18.2	34.3	84.5	24.1	584	*O. ficus-indica*
PARL 584-3-6	1303.6	19.3	43.2	99.6	23.7	584	*O. ficus-indica*
PARL 584-3-7	960.3	17.4	27.2	74.5	30.7	584	*O. ficus-indica*
PARL 584-3-8	760.9	15.6	30.5	73.5	24.4	584	*O. ficus-indica*
PARL 584-4-1	1059.3	18.9	35.2	86.3	21.6	584	*O. ficus-indica*
PARL 584-4-2	993.3	19.8	30.7	81.9	23	584	*O. ficus-indica*
PARL 584-4-3	584.2	15.6	24.3	66.4	21.6	584	*O. ficus-indica*
PARL 584-4-4	1174.3	19.9	40.3	95.9	20.9	584	*O. ficus-indica*
PARL 584-4-5	1511.2	19	39.2	96.9	29.9	584	*O. ficus-indica*
PARL 584-4-6	1529.1	23.8	42.7	103.7	20.3	584	*O. ficus-indica*
PARL 584-4-7	720.2	16.3	34.8	81.4	17.7	584	*O. ficus-indica*
PARL 584-4-8	916	17	38.3	90.4	19.8	584	*O. ficus-indica*
PARL 585-1-1	956.2	17.5	35.4	83.5	22.45	585	*O. ficus-indica*
PARL 585-1-2	851.6	17.7	32.5	80.1	24.5	585	*O. ficus-indica*
PARL 585-1-3	1212.4	22.6	38.6	95.2	22	585	*O. ficus-indica*
PARL 585-1-4	439.9	14.5	26	64.5	19.5	585	*O. ficus-indica*
PARL 585-1-5	382.1	16.2	27.8	65.4	15	585	*O. ficus-indica*
PARL 585-1-6	996.1	18.7	40.8	93.9	26	585	*O. ficus-indica*
PARL 585-1-7	722.3	15.8	34.5	83	21	585	*O. ficus-indica*
PARL 585-1-8	655.8	15.2	29	23.5	24	585	*O. ficus-indica*
PARL 585-2-1	787.4	18.4	29.9	75.4	20	585	*O. ficus-indica*
PARL 585-2-2	880.5	20.2	32.6	83.1	20.9	585	*O. ficus-indica*
PARL 585-2-3	457.2	18.3	33	72.6	15.3	585	*O. ficus-indica*
PARL 585-2-4	973.9	19.3	35.9	83	21.4	585	*O. ficus-indica*
PARL 585-2-5	581.2	16.2	28.5	71.9	18.2	585	*O. ficus-indica*
PARL 585-2-6	829.8	17.4	32.4	81.3	20.8	585	*O. ficus-indica*
PARL 585-2-7	452.4	14.2	22.9	60.4	20.1	585	*O. ficus-indica*
PARL 585-2-8	692	18.3	33	69.8	18.5	585	*O. ficus-indica*
PARL 585-3-1	1020.1	23.3	37.9	88.4	19.6	585	*O. ficus-indica*
PARL 585-3-2	981.4	21.3	38.6	90.7	20.4	585	*O. ficus-indica*
PARL 585-3-3	918.4	20.6	32.4	80.6	21.6	585	*O. ficus-indica*
PARL 585-3-4	818	20.5	36.6	85.8	21.8	585	*O. ficus-indica*

PARL 585-3-5	449	16.4	32.8	76.7	14.4	585	*O. ficus-indica*
PARL 585-3-6	903.4	19	33.1	82.8	25.2	585	*O. ficus-indica*
PARL 585-3-7	963.8	19.9	30.7	77.6	26.2	585	*O. ficus-indica*
PARL 585-3-8	1175.2	20.2	35.1	85.1	26.3	585	*O. ficus-indica*
PARL 585-4-1	649.8	16.6	33.6	81.4	16.5	585	*O. ficus-indica*
PARL 585-4-2	828.5	18.6	32.7	79.6	22.2	585	*O. ficus-indica*
PARL 585-4-3	544.2	15.4	22.6	57.8	26.6	585	*O. ficus-indica*
PARL 585-4-4	445.9	16.5	29.1	66.8	15	585	*O. ficus-indica*
PARL 585-4-5	860.9	18.5	34.4	82	23.3	585	*O. ficus-indica*
PARL 585-4-6	1015.2	18.6	36.7	86.1	22.6	585	*O. ficus-indica*
PARL 585-4-7	678	21	30.9	75.1	21.6	585	*O. ficus-indica*
PARL 585-4-8	762.8	17.9	32.1	79	23.1	585	*O. ficus-indica*
PARL 839-1-1	1587.7	29.4	46.9	119.4	17	839	*O. undulata*
PARL 839-1-2	941.3	35	41	95.3	14	839	*O. undulata*
PARL 839-1-3	885.9	24.6	34.7	94.6	19	839	*O. undulata*
PARL 839-1-4	1131.2	24.2	34.5	94.5	22	839	*O. undulata*
PARL 839-1-5	1147.4	31.1	37.4	99.4	15	839	*O. undulata*
PARL 839-1-6	948	24.4	31	90.5	17	839	*O. undulata*
PARL 839-1-7	1267	25.6	37.8	99.7	16	839	*O. undulata*
PARL 839-1-8	1098.2	23.4	38.4	98.5	17	839	*O. undulata*
PARL 839-2-1	557.3	20.3	30.1	74.9	17.9	839	*O. undulata*
PARL 839-2-2	939.9	21.9	35.2	88.2	21.5	839	*O. undulata*
PARL 839-2-3	361.9	15.6	26	66.5	15.2	839	*O. undulata*
PARL 839-2-4	918.8	23.9	34.2	94	12.5	839	*O. undulata*
PARL 839-2-5	1362.3	28	38.7	107.2	16.2	839	*O. undulata*
PARL 839-2-6	787.8	24.1	31.9	88.9	16.4	839	*O. undulata*
PARL 839-2-7	703.6	17	26	73.4	28.7	839	*O. undulata*
PARL 839-2-8	1254.7	26.4	36.9	102.5	18.4	839	*O. undulata*
PARL 839-3-1	1009.4	23.9	35.5	87.2	21.5	839	*O. undulata*
PARL 839-3-2	946.4	27.9	41.4	97.9	11.8	839	*O. undulata*
PARL 839-3-3	987	24.7	36.1	96.5	14.6	839	*O. undulata*
PARL 839-3-4	877.6	24.4	31.4	89.9	16	839	*O. undulata*
PARL 839-3-5	1263.5	26.1	34	95.1	23.1	839	*O. undulata*
PARL 839-3-6	943.8	23.6	39.2	98.4	15	839	*O. undulata*
PARL 839-3-7	946	22.6	34	89.6	20.2	839	*O. undulata*
PARL 839-3-8	900.4	22.1	40.4	96.3	17.5	839	*O. undulata*
PARL 839-4-1	996.7	25.6	36.8	101.5	14.8	839	*O. undulata*
PARL 839-4-2	788.7	23.7	36.9	90.3	12.4	839	*O. undulata*
PARL 839-4-3	599	21.6	36.6	85.8	11.2	839	*O. undulata*
PARL 839-4-4	1481.9	28.6	36.9	106.9	18.8	839	*O. undulata*
PARL 839-4-5	417.8	18.8	31.7	75.2	9.97	839	*O. undulata*
PARL 839-4-6	898.3	28.2	30.6	88.6	15	839	*O. undulata*
PARL 839-4-7	948.9	22.4	42.3	101.6	11.5	839	*O. undulata*

PARL 839-4-8	674.9	22.8	23.9	79.3	20.4	839	*O.undulata*
PARL 845-1-1	857	22.6	35.5	92.5	14.3	845	*O. ficus-indica*
PARL 845-1-2	454.5	17.1	31.6	75.4	13.5	845	*O. ficus-indica*
PARL 845-1-3	831	22.4	28.9	79.3	19.3	845	*O. ficus-indica*
PARL 845-1-4	734.8	21.8	29.9	75	18.3	845	*O. ficus-indica*
PARL 845-1-5	464.2	15.8	24.3	66.2	16.3	845	*O. ficus-indica*
PARL 845-1-6	719.7	19.2	26.9	73.7	18.3	845	*O. ficus-indica*
PARL 845-1-7	958.2	21.3	33.6	87.6	19.4	845	*O. ficus-indica*
PARL 845-1-8	830.7	25.5	35.4	89	13.3	845	*O. ficus-indica*
PARL 845-2-1	457.8	15.2	26.2	63.7	19.2	845	*O. ficus-indica*
PARL 845-2-2	1121.6	24.6	35.2	93.9	17.3	845	*O. ficus-indica*
PARL 845-2-3	1398.4	23.5	41.6	103.4	15	845	*O. ficus-indica*
PARL 845-2-4	744.3	20.1	31.6	82.3	12.9	845	*O. ficus-indica*
PARL 845-2-5	557	17.6	24.5	68	17.6	845	*O. ficus-indica*
PARL 845-2-6	373.7	19.4	23.2	62.3	11.1	845	*O. ficus-indica*
PARL 845-2-7	1596	25.3	42.5	110	21.3	845	*O. ficus-indica*
PARL 845-2-8	1067.1	19.6	36.2	92.7	16.7	845	*O. ficus-indica*
PARL 845-3-1	1213.5	24.2	35.4	92.1	21.2	845	*O. ficus-indica*
PARL 845-3-2	699.4	21.3	33.9	86.5	16.8	845	*O. ficus-indica*
PARL 845-3-3	909.4	20.7	28.1	77.1	27.4	845	*O. ficus-indica*
PARL 845-3-4	1207.1	22.9	38.4	97.9	19.7	845	*O. ficus-indica*
PARL 845-3-5	555.4	19.2	27.5	73.4	12.2	845	*O. ficus-indica*
PARL 845-3-6	888.6	22.3	34.8	89.6	16.9	845	*O. ficus-indica*
PARL 845-3-7	1191.6	22.4	37.2	94.8	20.6	845	*O. ficus-indica*
PARL 845-3-8	1355.9	25.1	43.2	106.4	19.2	845	*O. ficus-indica*
PARL 845-4-1	456.6	15.7	25.6	66.3	16.2	845	*O. ficus-indica*
PARL 845-4-2	829.5	19.6	33.9	81.8	21.4	845	*O. ficus-indica*
PARL 845-4-3	749.3	20.2	33.4	86.6	18.3	845	*O. ficus-indica*
PARL 845-4-4	501.5	16.1	29.1	65.7	20.3	845	*O. ficus-indica*
PARL 845-4-5	495.5	17.1	24.2	64.1	19.5	845	*O. ficus-indica*
PARL 845-4-6	609.8	18.3	29.7	75.3	16.8	845	*O. ficus-indica*
PARL 845-4-7	660.4	17.3	34.6	84.2	16.8	845	*O. ficus-indica*
PARL 845-4-8	467.9	18.2	27.4	69.2	16.9	845	*O. ficus-indica*
PARL 854-1-1	645.5	18.6	28	73.6	23.41	854	*O. ficus-indica*
PARL 854-1-2	479.6	15.3	24	64	20.47	854	*O. ficus-indica*
PARL 854-1-3	436.7	16.2	30.6	72.4	21.27	854	*O. ficus-indica*
PARL 854-1-4	307.6	14.2	24	62.5	13	854	*O. ficus-indica*
PARL 854-1-5	564.7	19.3	28.4	71.7	18.34	854	*O. ficus-indica*
PARL 854-1-6	517.5	18	27	69.5	16.27	854	*O. ficus-indica*
PARL 854-1-7	578.7	17.1	25.6	67.7	21.54	854	*O. ficus-indica*
PARL 854-1-8	398.5	14.5	24.5	62.5	17.32	854	*O. ficus-indica*
PARL 854-2-1	489	16.4	27.1	65.7	19.4	854	*O. ficus-indica*
PARL 854-2-2	319.8	13.3	19.5	51.8	22.5	854	*O. ficus-indica*

PARL 854-2-3	448.4	14.6	31.1	30.6	16.3	854	*O. ficus-indica*
PARL 854-2-4	347.5	15.7	25.5	62.2	14.6	854	*O. ficus-indica*
PARL 854-2-5	428	18.7	30.1	73.2	13.4	854	*O. ficus-indica*
PARL 854-2-6	377.9	13.8	25.9	65.7	16.6	854	*O. ficus-indica*
PARL 854-2-7	323.8	13.4	22.5	53.2	18.4	854	*O. ficus-indica*
PARL 854-2-8	365.4	14.5	27.4	65.5	16.6	854	*O. ficus-indica*
PARL 854-3-1	448.2	17.3	29.1	70.6	14.6	854	*O. ficus-indica*
PARL 854-3-2	379.7	15.5	23.9	62	16.3	854	*O. ficus-indica*
PARL 854-3-3	428.9	14.9	28.8	66.8	18	854	*O. ficus-indica*
PARL 854-3-4	491	16.6	28	70.1	18.2	854	*O. ficus-indica*
PARL 854-3-5	394.9	15.9	22	59	20.6	854	*O. ficus-indica*
PARL 854-3-6	490.5	16.5	24.2	62.4	19.6	854	*O. ficus-indica*
PARL 854-3-7	348.2	14.5	20.4	54.5	17.4	854	*O. ficus-indica*
PARL 854-3-8	385.6	17.1	30.8	68.5	11.8	854	*O. ficus-indica*
PARL 854-4-1	258.5	13.2	23.9	58.1	13.6	854	*O. ficus-indica*
PARL 854-4-2	407.3	14.3	24.9	59.1	20.9	854	*O. ficus-indica*
PARL 854-4-3	355.9	15.6	20.8	55.4	21.3	854	*O. ficus-indica*
PARL 854-4-4	484.8	16.3	22.7	58.4	22.3	854	*O. ficus-indica*
PARL 854-4-5	580	16.4	28.6	69.4	16.2	854	*O. ficus-indica*
PARL 854-4-6	580.2	18.5	24.4	66.4	19.3	854	*O. ficus-indica*
PARL 854-4-7	301.7	13.3	23.1	58.7	15.7	854	*O. ficus-indica*
PARL 854-4-8	414.8	14.8	22.6	56.9	21.6	854	*O. ficus-indica*

A3. All violin plots on the unfiltered (left) and filtered (right) datasets on the initial measures, organized by accession.
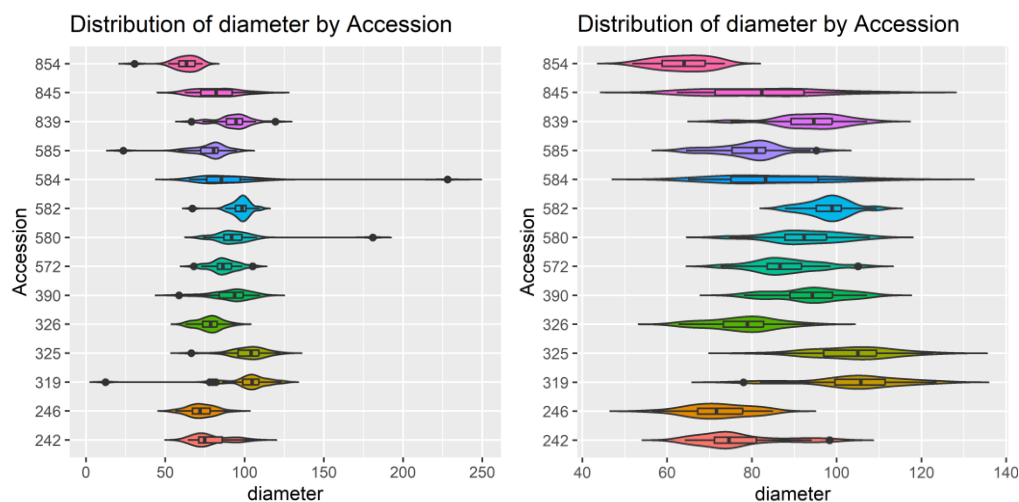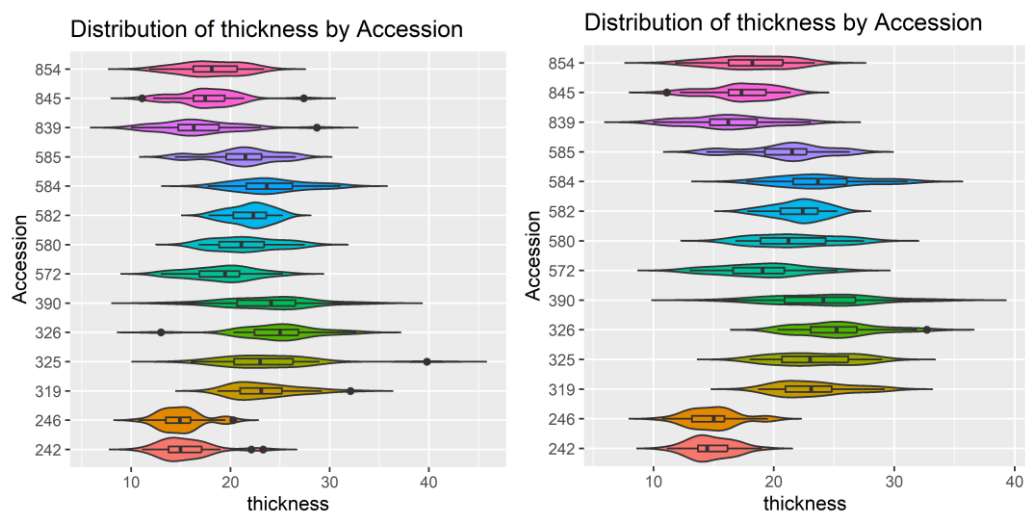

A3.1. A violin plot of diameter value distributions for each accession generated from the unfiltered (left) and filtered (right) datasets on the initial measures, organized by accession. Outliers (black dots) are highlighted. Interquartile ranges have been included in each violin shape as rectangles within the violin shapes.
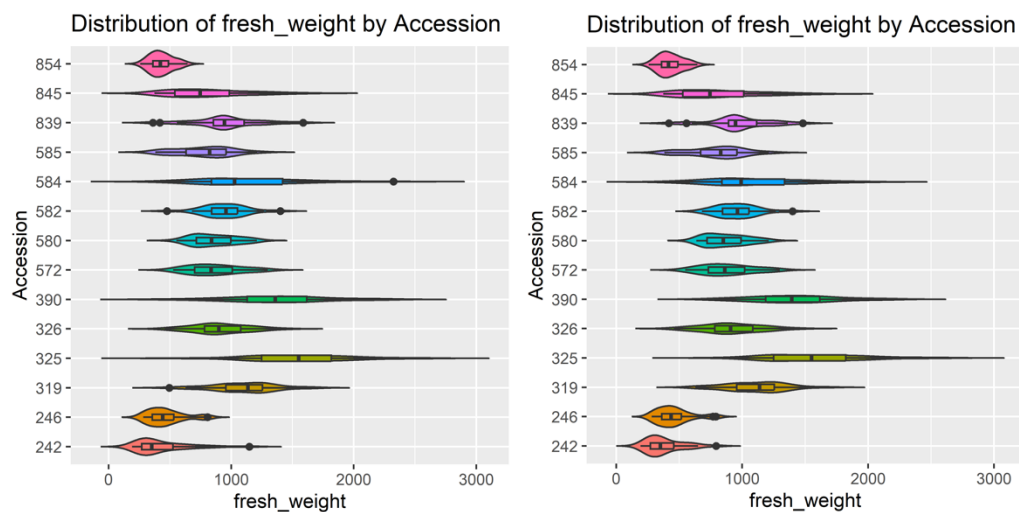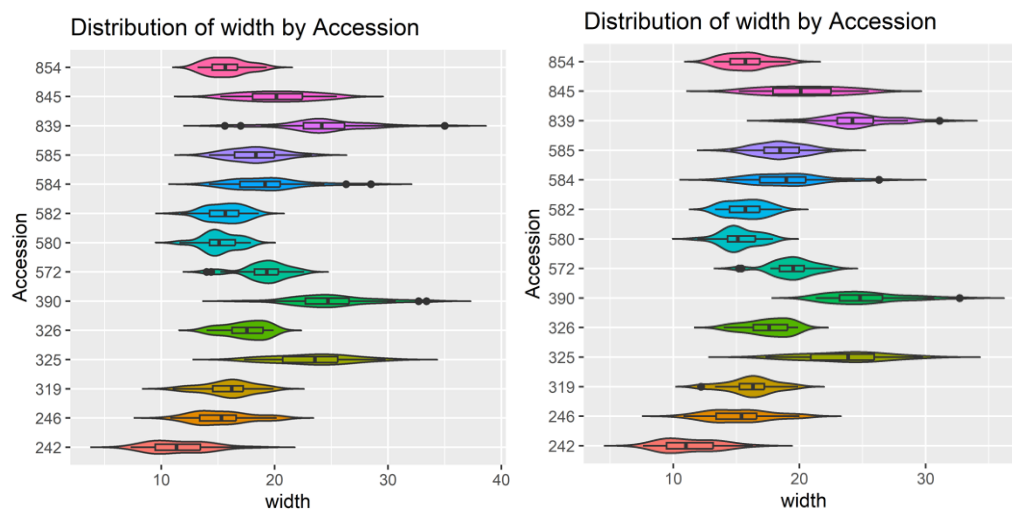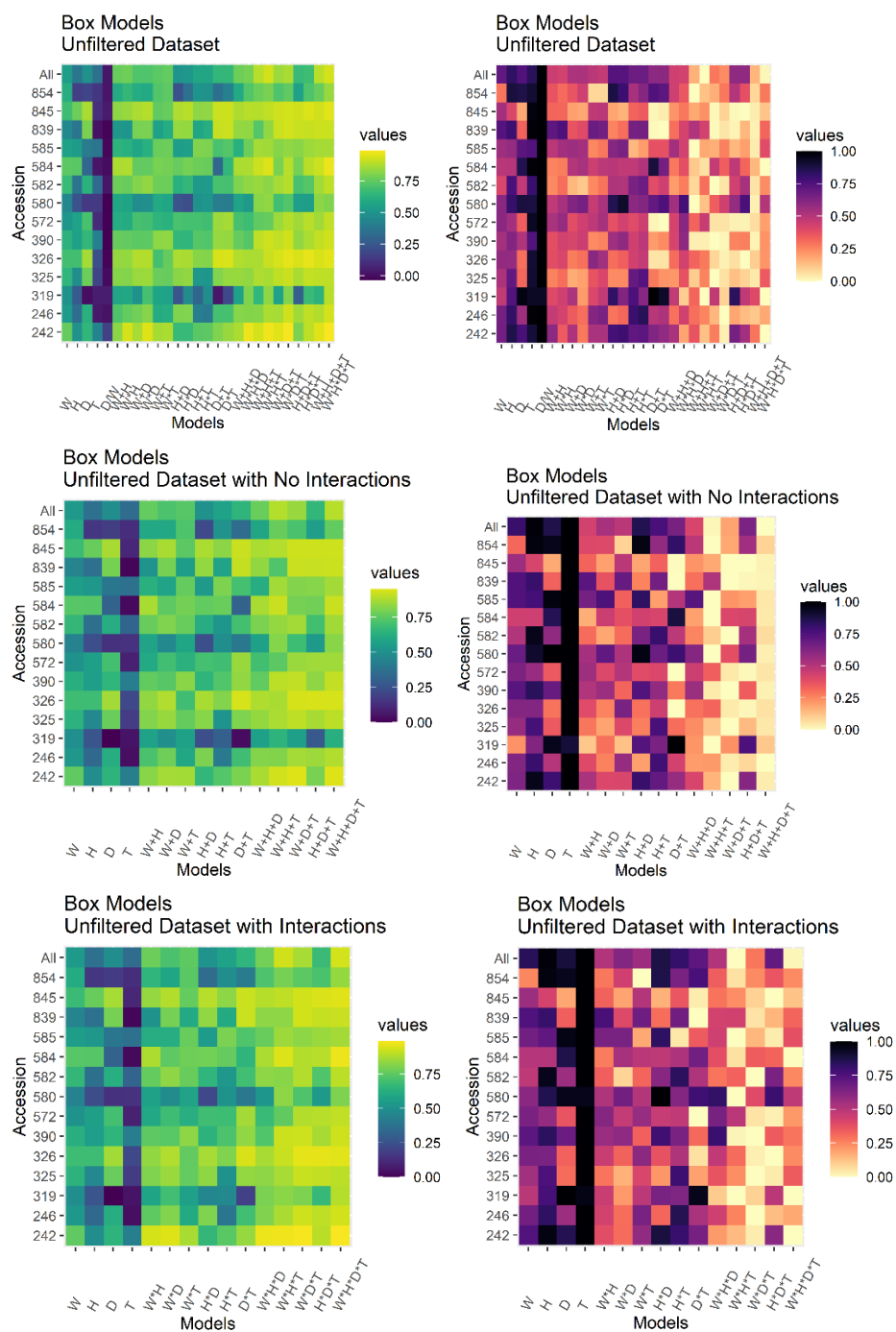

A3.2. A violin plot of fresh weight value distributions for each accession generated from the unfiltered (left) and filtered (right) datasets on the initial measures, organized by accession. Outliers (black dots) are highlighted. Interquartile ranges have been included in each violin shape as rectangles within the violin shapes.
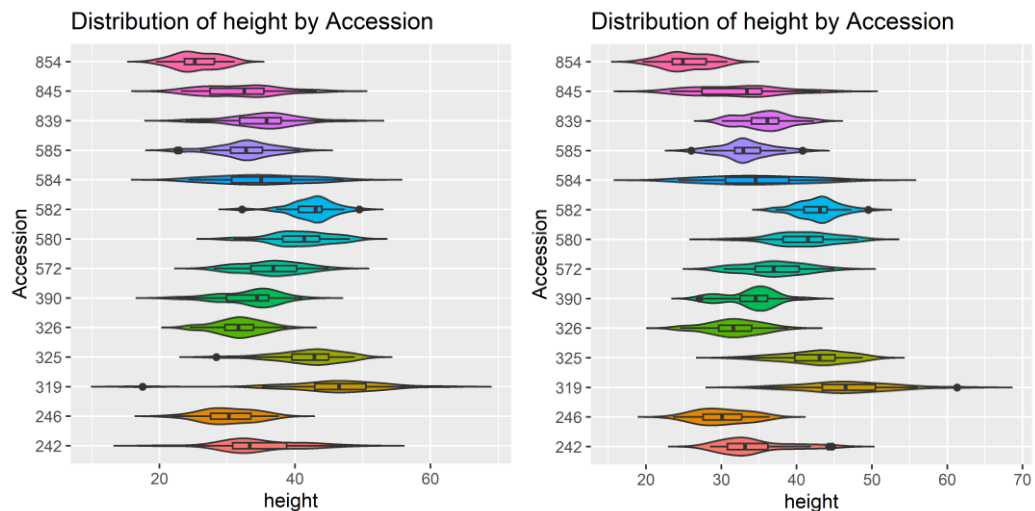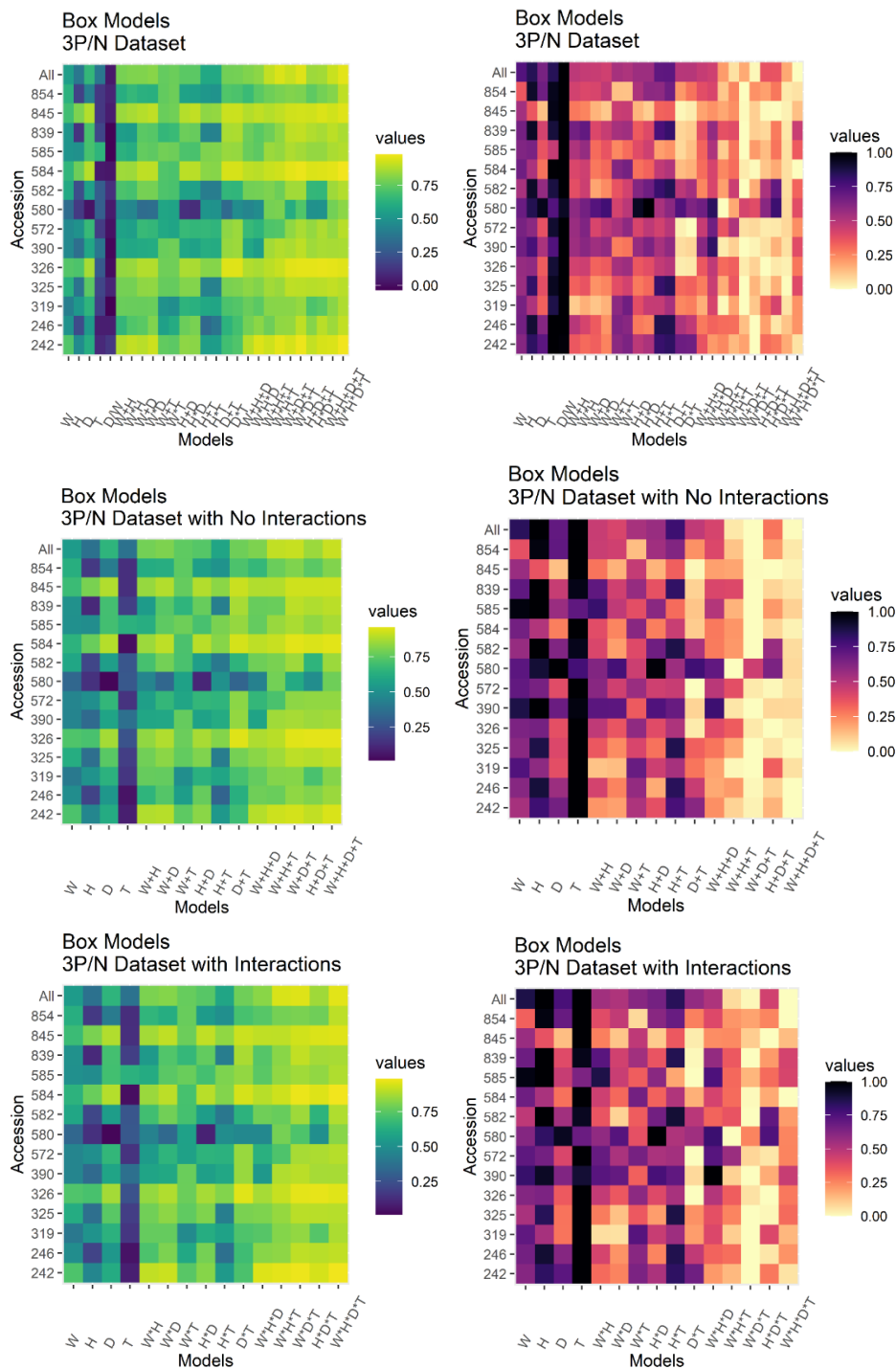
A3.3. A violin plot of height value distributions for each accession generated from the unfiltered (left) and filtered (right) datasets on the initial measures, organized by accession. Outliers (black dots) are highlighted. Interquartile ranges have been included in each violin shape as rectangles within the violin shapes.
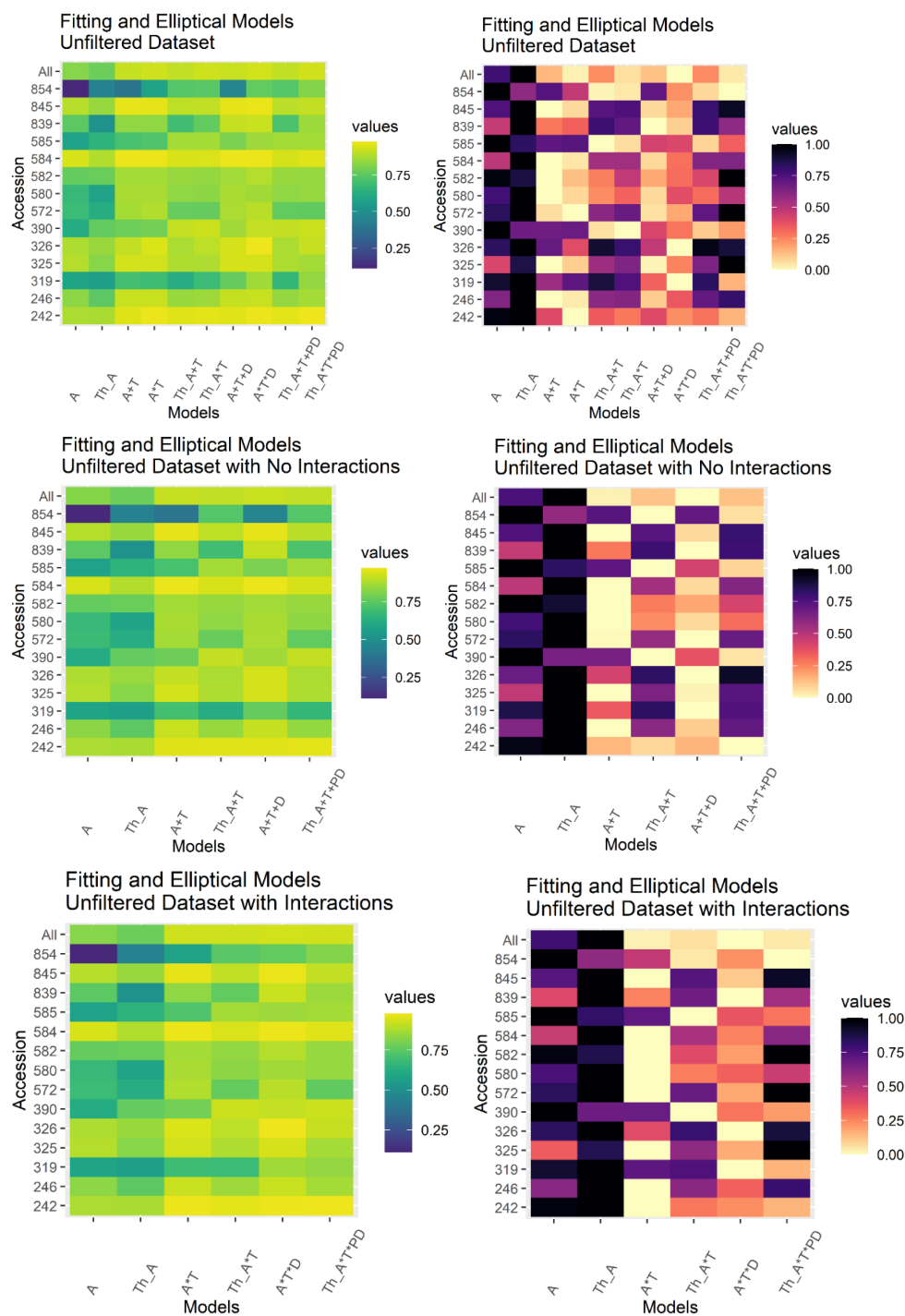

A3.4. A violin plot of thickness value distributions for each accession generated from the unfiltered (left) and filtered (right) datasets on the initial measures, organized by accession. Outliers (black dots) are highlighted. Interquartile ranges have been included in each violin shape as rectangles within the violin shapes.

A3.5. A violin plot of width value distributions for each accession generated from the unfiltered (left) and filtered (right) datasets on the initial measures, organized by accession. Outliers (black dots) are highlighted. Interquartile ranges have been included in each violin shape as rectangles within the violin shapes.

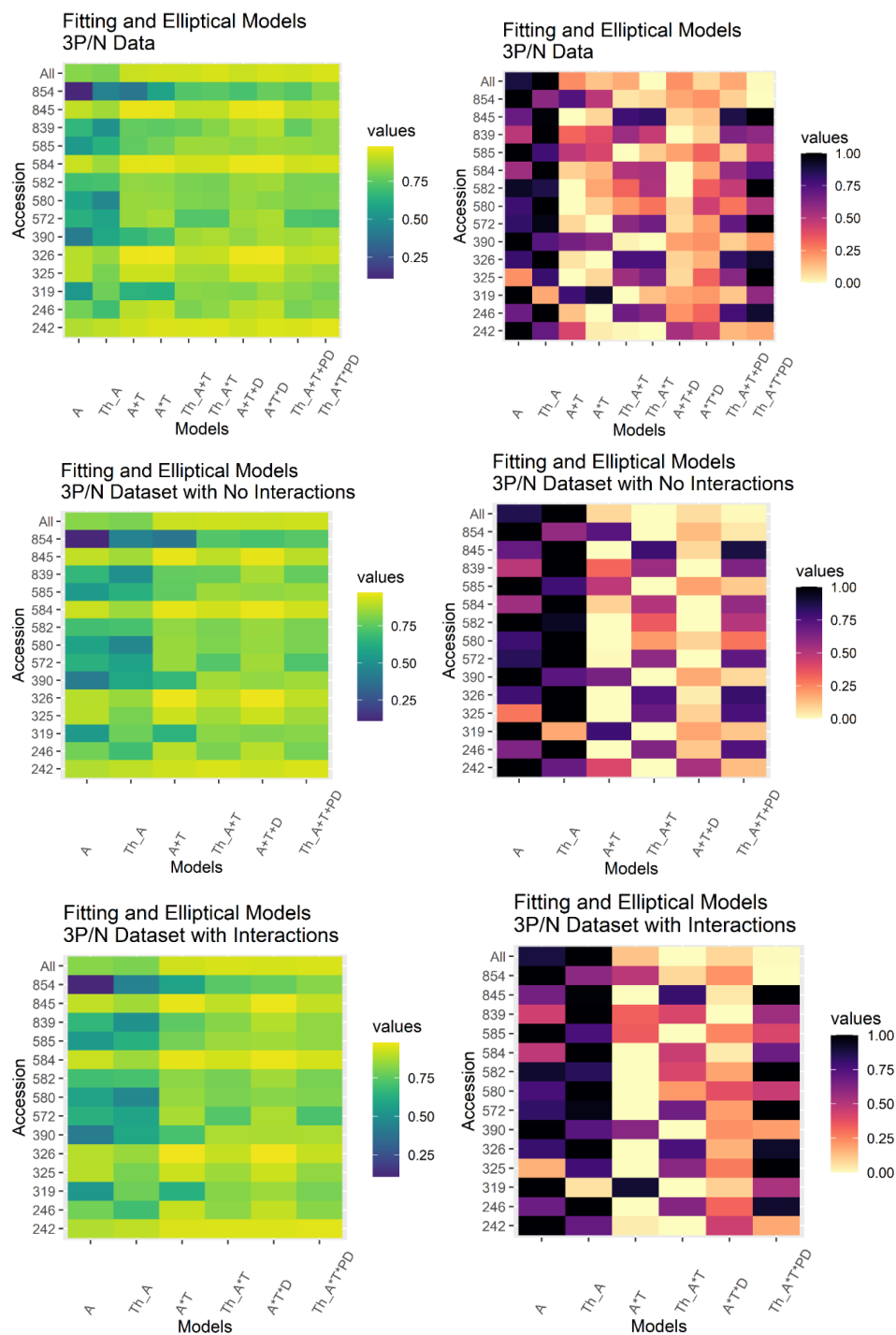**A4.** Expanded model testing heatmaps generated over filtered/unfiltered data

A4.1a-f. A heatmap plot of all possible combinations of the four base measures for the box model, evaluated using an adjusted R^2 value (figures a, c, e) and normalized SBC criterion value (b, d, f) over the *unfiltered* dataset. For comparison, plots have been generated with all possible models (a, b), models without interaction terms as denoted by + (c, d) and models with interaction terms included as denoted by the * operator (e, f). Sections a, c, and e evaluate adjusted R^2 fit using a viridis color palette where blue is a lower value and yellow is a higher, more optimal value. Sections b, d, and f evaluate normalized SBC values using a viridis color palette where the lighter color indicates the lowest and most optimal SBC value for the row, while the darker color indicates the highest value and least parsimonious.

A4.2a-f. A heatmap plot of all possible combinations of the four base measures for the box model, evaluated using an adjusted R^2 value (figures a, c, e) and normalized SBC criterion value (b, d, f) over the *filtered* dataset. For comparison, plots have been generated with all possible models (a, b), models without interaction terms as denoted by "+" (c, d) and models with interaction terms included as denoted by the * operator (e, f). Sections a, c, and e evaluate adjusted R^2 fit using a viridis color palette where blue is a lower value and yellow is a higher, more optimal value. Sections b, d, and f evaluate normalized SBC values using a viridis color palette where the lighter color indicates the lowest and most optimal SBC value for the row, while the darker color indicates the highest value and least parsimonious.
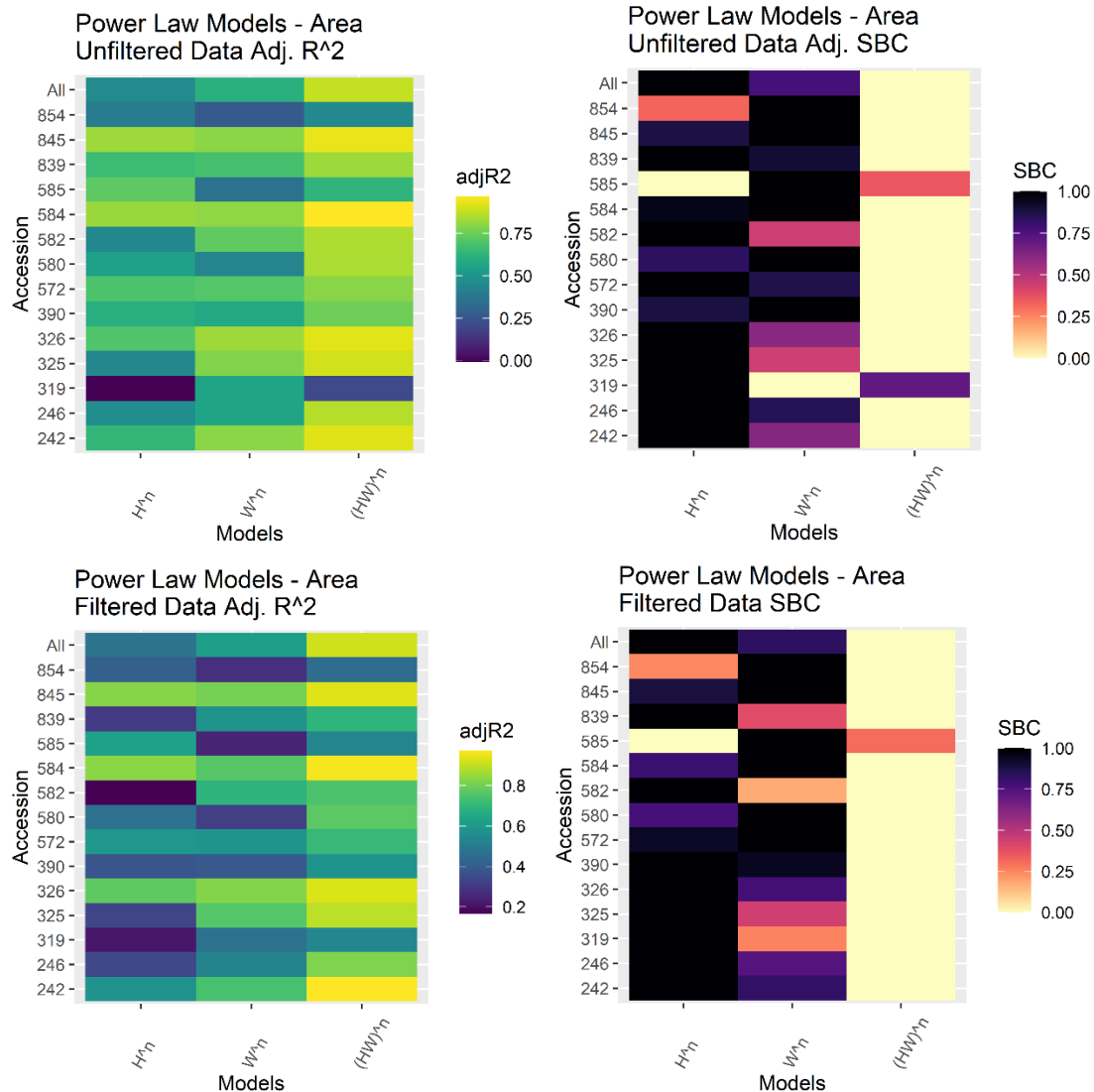
A4.3a-f. A heatmap plot of all possible combinations of area and diameter measures for the fitting box and elliptical models, evaluated using an adjusted R^2 value (figures a, c, e) and normalized SBC criterion value (b, d, f) over the *unfiltered* dataset. For comparison, plots have been generated with all possible models (a, b), models without interaction terms as denoted by + (c, d) and models with interaction terms included as denoted by the * operator (e, f). Sections a, c, and e evaluate adjusted R^2 fit using a viridis color palette where blue is a lower value and yellow is a higher, more optimal value. Sections b, d, and f evaluate normalized SBC values using a viridis color palette where the lighter color indicates the lowest and most optimal SBC value for the row, while the darker color indicates the highest value and least parsimonious.
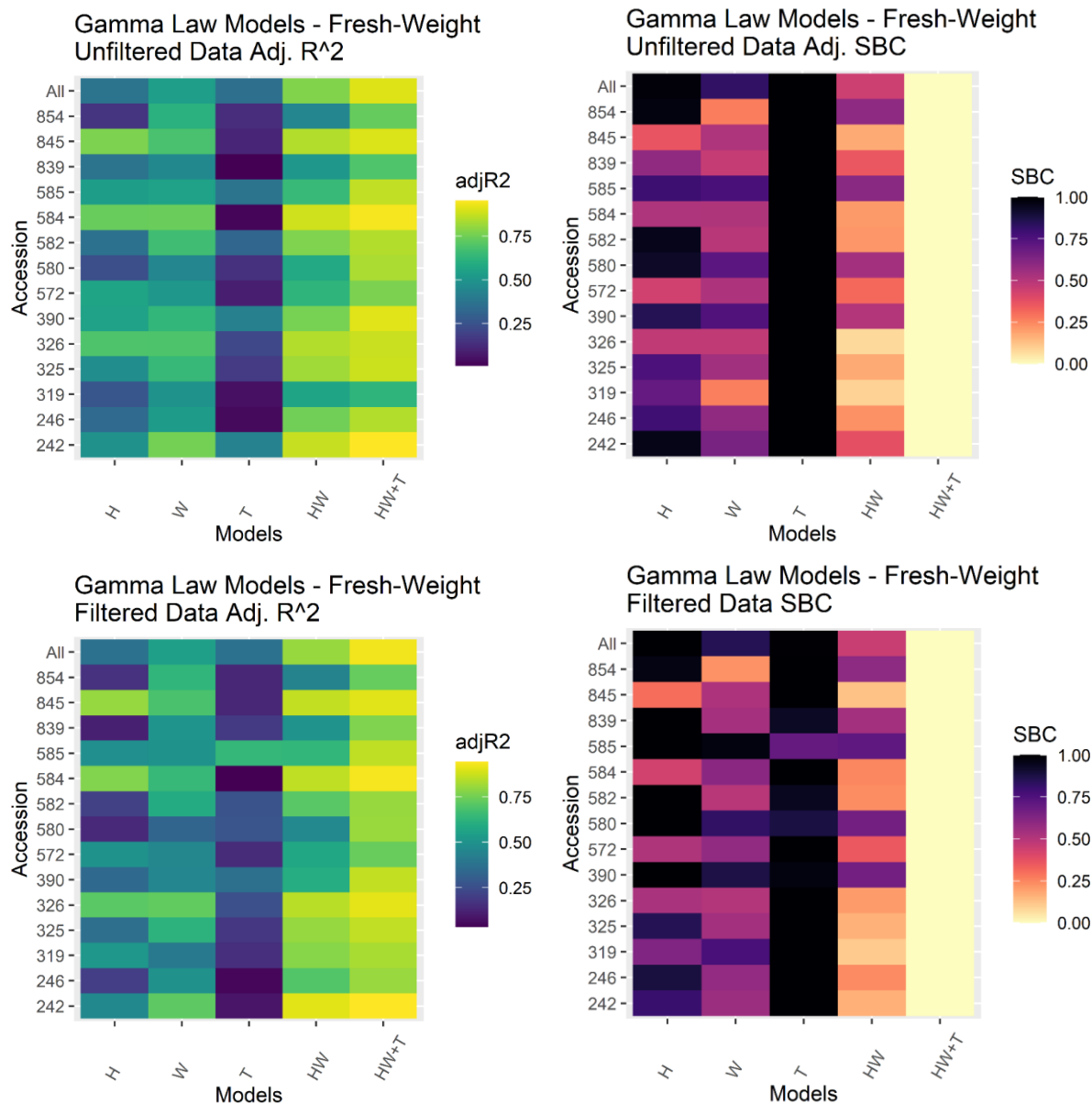
A4.4a-f. A heatmap plot of all possible combinations of area and diameter measures for the fitting box and elliptical models, evaluated using an adjusted R^2 value (figures a, c, e) and normalized SBC criterion value (b, d, f) over the *filtered* dataset. For comparison, plots have been generated with all possible models (a, b), models without interaction terms as denoted by + (c, d) and models with interaction terms included as denoted by the * operator (e, f). Sections a, c, and e evaluate adjusted R^2 fit using a viridis color palette where blue is a lower value and yellow is a higher, more optimal value. Sections b, d, and f evaluate normalized SBC values using a viridis color palette where the lighter color indicates the lowest and most optimal SBC value for the row, while the darker color indicates the highest value and least parsimonious.

A4.5a-d. A heatmap plot of all possible combinations of three base measures for the power law models generated for cladode area in Lucena et. al. 2021, evaluated using an adjusted R^2 value (figures a and c) and normalized SBC criterion value (b and d) over both the *unfiltered* (a and b) and *filtered* (c and d) dataset. Sections a and c evaluate adjusted R^2 fit using a viridis color palette where blue is a lower value and yellow is a higher, more optimal value. Sections b and d evaluate normalized SBC values using a viridis color palette where the lighter color indicates the lowest and most optimal SBC value for the row, while the darker color indicates the highest value and least parsimonious.

A4.6a-d. A heatmap plot of all possible combinations of three base measures for the gamma law models generated for cladode fresh weight in Lucena et. al. 2021, evaluated using an adjusted R^2 value (figures a and c) and normalized SBC criterion value (b and d) over both the *unfiltered* (a and b) and *filtered* (c and d) dataset. Sections a and c evaluate adjusted R^2 fit using a viridis color palette where blue is a lower value and yellow is a higher, more optimal value. Sections b and d evaluate normalized SBC values using a viridis color palette where the lighter color indicates the lowest and most optimal SBC value for the row, while the darker color indicates the highest value and least parsimonious.

A5. R Program Code

A5.1 – RunAll.R

```
# These scripts were written to automate on a large scale the statistical
analyses
# done in R, in order to save time - so that minor changes to analytical
processes
# may be quickly reflected across the entire analysis. This allows hours of
work
# to be repeated, with or without changes, in less than two minutes while
analyzing
# thousands of datapoints.
#
# This also permits the data analysis process to be inspected by reviewers at
each
# step of the analysis process.
#
# The program is interspersed with diagnostic output messages ("cat(...)")
that
# output to the console and indicate the progression of the program.
#
# Contact gangres@nevada.unr.edu for questions regarding the code's function
and
# behavior.


#This script runs four custom function libraries, a file-loading script, and
a file- and figure-generating in order.
#UTIL, GRAPHER, COMPILE, and CALC must be run first, in no particular order.
#Then, "File Loading and Prep" and "File Generation" must be run, in that
specific order.


# |} Establish working directory with target files.
# Workstation Setup ------------------------------------------------------

STARTTIME <- Sys.time()

cat("Setting Working Directory...\n")

#REMEMBER TO OMIT THIS PRIOR TO PROGRAM SUBMISSION
setwd("C:\\Users\\gjang\\Documents\\GitHub\\CAI-Analysis-Script\\Script-v2")

cat("Locating Dataset File...\n")
DATA_FILE <- 'PARLO_06202022.csv'

cat("-------------------------------\n")
cat("-------------------------------\n")
cat("LOADING UTILITY FUNCTIONS")
cat("\n")
source("UTIL-Functions.R")
cat("\n")

cat("-------------------------------\n")
cat("-------------------------------\n")
cat("LOADING CALCULATOR FUNCTIONS\n")
cat("\n")
source("CALC-Functions.R")
cat("\n")

cat("-------------------------------\n")
cat("-------------------------------\n")
cat("LOADING GRAPHER FUNCTIONS\n")
```

```
cat("\n")
source("GRAPHER-Functions.R")
cat("\n")

cat("--------------------------------\n")
cat("--------------------------------\n")
cat("LOADING COMPILATION FUNTIONS\n")
cat("\n")
source("COMPILE-Functions.R")
cat("\n")

cat("--------------------------------\n")
cat("--------------------------------\n")
cat("CREATING FILES\n")
cat("\n")
source("File Loading and Prep.R")
cat("\n")

cat("--------------------------------\n")
cat("--------------------------------\n")
cat("CREATING OUTPUT GRAPHS\n")
cat("\n")
source("File Generation.R")
cat("\n")

ENDTIME <- Sys.time()
TOTALTIME <- ENDTIME - STARTTIME

cat("Total Time Elapsed: ")
cat(TOTALTIME)
cat(" Minutes")
```

A5.2 – UTIL-Functions.R

```
#' @name UTIL.toModelString
#' @description Takes two strings and assembles them into a model name
parseable by an lm() call.
#' @param xname A string denoting the independent variable.
#' @param yname A string denoting the denpendent variable.
#' @example UTIL.toModelString("width", "height") >> "height~width"
#' @returns A string containing a modelname.

UTIL.toModelString <- function(xname, yname) {
  return(paste0(yname, "~", xname))
}

# Filtration Utilities ------------------------------------------------

#' @name UTIL.AugmentModelWithID
#' @description Adds IDs for each datapoint with a given accession number.
#' @param model_in The linear model object to be augmented.
#' @param accession A string containing the accession ID for the
augmentation.
#' @returns An augmented linear model object with a ID column within the
model_obj$model property.

UTIL.AugmentModelWithID <- function(model_in, accession) {
  model_out <- model_in
  ID_vector <- c()
  ID_array <- array(c(1:32), dim = c(8, 4))
  ID_rows <- row(ID_array)
  ID_cols <- col(ID_array)
  for (n in c(1:32)) {
    ID_1 <- ID_cols[n]
    ID_2 <- ID_rows[n]
    ID_vector <- c(ID_vector, paste0("PARL ", accession, "-", ID_1, "-",
ID_2))
  }
  model_out$model$ID <- ID_vector
  return(model_out)
}

#' @name Assemble3PNDroplist
#' @description Collects all datapoints which exceed the 3P/N leverage
threshold in the models selected, by accession.
#' @param fits_df A dataframe containing pregenerated models from
CALC.ModelGenerator.
#' @param data_df A dataframe containing all measured data points with IDs
and their measures.
#' @returns A vector with string IDs.

UTIL.Assemble3PNDroplist <- function(fits_df, data_df) {
  blacklist_3pn <- c()

  for (modelname in colnames(fits_df)[-1]) {

    for (accession in SWITCHBOARD.ACCESSIONLIST) {

      model <- CALC.ModelFetch(fits_df, accession, modelname)
      target_loc <- CALC.ModelIndexHighLeverage(model)

      for (loc in target_loc) {
        pairframe <- CALC.ModelValuePairsFromIndex(loc, model)
        modelID <- CALC.PadIDfromModelValuePairs(pairframe, data_df,
accession)
        blacklist_3pn <- c(blacklist_3pn, modelID)
```

```
        }
      }
    }
    blacklist_3pn <- sort(unique(blacklist_3pn))
    return(blacklist_3pn)
}

#' @name UTIL.DropOnNPN
#' @description Removes IDs in a droplist from a larger dataset.
#' @param data_df A dataframe containing all measured data points with IDs
and their measures.
#' @param droplist_NPN A vector of strings containing IDs of the datapoints
to drop.
#' @returns A dataframe excluding the removed datapoints.

UTIL.DropOnNPN <- function(data_df, droplist_NPN) {
    data_out <- data_df[-which(data_df$completeID %in% droplist_NPN),]
    #print(nrow(data_out))
    return(data_out)
}

#' @name UTIL.AugmentModelWithLeverageTagging
#' @description Adds data on whether or not each datapoint exceeds the 3P/N
leverage threshold for the particular model.
#' @param model_in The linear model object to be augmented.
#' @returns An augmented linear model object with a tagged column within the
model_obj$model property.

UTIL.AugmentModelWithLeverageTagging <- function(model_in) {
    model_out <- model_in

    indices_HI <- model_out %>%
        CALC.ModelIndexHighLeverage()

    model_out$model$tagged <- "N/A"
    model_out$model$tagged[indices_HI] <- ">3p/n Threshold"

    return(model_out)
}

#' @name UTIL.AllAccessionToMain
#' @description Appends a row of summary statistic values (for models applied
over the entire dataset) to a dataframe containing the by-accession table of
values.
#' @param model_df The summary statistic dataframe to be augmented.
#' @param add_list A list of values to be appended to the dataframe.
#' @returns A dataframe with an additional row and the same amount of
columns.

UTIL.AllAccessionToMain <- function(model_df, add_list) {
    rowaddloc <- nrow(model_df) + 1
    model_df[rowaddloc, 1] <- "All"
    for (model in names(add_list)) {
        coladdloc <- which(names(add_list) == model)
        model_df[rowaddloc, coladdloc+1] <- add_list[[model]]
    }

    return(model_df)
}


# File Creation Utilities --------------------------------------------------

#' @name UTIL.FolderPathCheck
```

```
#' @description Checks to see if a folder exists, and makes it if it doesn't.
#' @param subfolder_string A string or path containing a valid subfolder.

UTIL.FolderPathCheck <- function(subfolder_string) {
  path_components <- str_split(subfolder_string, "\\\\")[[1]]
  for (position in c(1:length(path_components))) {
    checkpath <- path_components[1:position] %>%
      paste0(collapse = "\\\\")
    if(!(file.exists(checkpath))) {
      dir.create(file.path(checkpath))
    }
  }
}

#' @name UTIL.quickPNG
#' @description Creates a .png object with a given name to store an image in.
#' @param filenamestring The name given for the .png file.
#' @param subfolder The folder that the file should be stored in.

UTIL.quickPNG <- function(filenamestring, subfolder, dimensions = c(600, 4,
4)) {

  png_ppi = dimensions[1]
  png_height = dimensions[2]
  png_width = dimensions[3]

  UTIL.FolderPathCheck(subfolder)
  filepath <- file.path(SWITCHBOARD.DIRECTORY, subfolder, filenamestring,
fsep = "\\")

  png(paste0(filepath, ".png"),
      width = png_width * png_ppi,
      height = png_height * png_ppi,
      res = png_ppi)
}

#' @name UTIL.quickCSV
#' @description Creates a .csv file from a dataframe and stores it in a
particular path.
#' @param data_df A dataframe containing all measured data points with IDs
and their measures.
#' @param filenamestring The name given for the .png file.
#' @param subfolder The folder that the file should be stored in.

UTIL.quickCSV <- function(data_df, filenamestring, subfolder) {

  UTIL.FolderPathCheck(subfolder)
  filepath <- file.path(SWITCHBOARD.DIRECTORY, subfolder, filenamestring,
fsep = "\\")

  write.csv(
    data_df,
    paste0(filepath, ".csv"),
    row.names = FALSE
  )
}

#' @name UTIL.StoreHeatmapsAndFile
#' @description Generates and stores a Heatmap image as a .png and its
corresponding table as a .csv file.
#' @param data_df A dataframe containing all measured data points with IDs
and their measures.
#' @param fill_models_df A vector of model strings.
#' @param filenamestring The name given for the .png file.
```

```
#' @param subfolder The folder that the file should be stored in.
#' @param abbreviations A vector of string abbreviations for the models for
intelligible graphing.

UTIL.StoreHeatmapsAndFile <- function(data_df, fill_models_df,
filenamestring, subfolder, heatmap_title, abbreviations) {

  if (!(dir.exists(subfolder))) {
    dir.create(subfolder)
  }

  all_R2 <- CALC.ModelGenerator_AllAccessions_R2(data_df, fill_models_df)
  all_SBC <- CALC.ModelGenerator_AllAccessions_SBC(data_df, fill_models_df)

  R2_frame_name <- paste0(filenamestring, "_R2")
  SBC_frame_name <- paste0(filenamestring, "_SBC")

  data_df %>%
    CALC.ModelGenerator(fill_models_df) %>%
    CALC.df_R2 %>%
    UTIL.AllAccessionToMain(all_R2) %>%
    `row.names<-`(NULL) %T>%
    COMPILE.R2heatmap(R2_frame_name, paste0(subfolder, "\\Graphs"),
heatmap_title, abbreviations) %>%
    UTIL.quickCSV(R2_frame_name, paste0(subfolder, "\\Tables"))

  data_df %>%
    CALC.ModelGenerator(fill_models_df) %>%
    CALC.df_SBC %>%
    UTIL.AllAccessionToMain(all_SBC) %>%
    `row.names<-`(NULL) %T>%
    COMPILE.SBCheatmap(SBC_frame_name, paste0(subfolder, "\\Graphs"),
heatmap_title, abbreviations) %>%
    UTIL.quickCSV(SBC_frame_name, paste0(subfolder, "\\Tables"))
}

#' @name UTIL.SplitModelListOnInteractions
#' @description Splits a master model list into two lists with separate + and
* models for more visually accessible heatmap graphing.
#' @param model_list A vector of model strings.
#' @param num_keep THe number of initial models to put in both output lists.
#' @returns A list with two string vectors.

UTIL.SplitModelListOnInteractions <- function(model_list, num_keep) {
  #print(model_list)
  keep_names <- model_list$models[1:num_keep]

  #no_interactions_piece <- model_list[which("+" %in% model_list$models),]
  #interactions_piece <- model_list[which("*" %in% model_list$models),]

  no_interactions_piece <- model_list[which(model_list$models %like% "[+]"),]
  interactions_piece <- model_list[which(model_list$models %like% "[*]"),]

  no_inter_list <- as.data.frame(c(keep_names, no_interactions_piece)) %>%
    `colnames<-`(c("models"))
  inter_list <- as.data.frame(c(keep_names, interactions_piece)) %>%
    `colnames<-`(c("models"))

  return(list(no_inter_list, inter_list))
}
```

A5.3 – CALC-Functions.R

```
# ModelGenerator Family -----------------------------------------------

#' @name CALC.ModelGenerator
#' @description Creates a dataframe of all possible linear regressions with
each permutation of model and accession.
#' @param data_df A dataframe containing all measured data points with IDs
and their measures.
#' @param models_df A dataframe containing the model strings for all models
of interest.
#' @returns A dataframe containing every linear model in \code{models_df},
generated for each accession in \code{SWITCHBOARD.ACCESSIONLIST}
#' @example CALC.ModelGenerator(SWITCHBOARD.csvMAINFILE,
SWITCHBOARD.MODELS_DF)

CALC.ModelGenerator <- function(data_df, fill_models_df) {

  fits_frame <- data.frame(accession = SWITCHBOARD.ACCESSIONLIST)
  for(model in fill_models_df$models) {
    new_models <- data_df %>%
      group_by(accession) %>%
      do(model = lm(model, data = .)) %>%
      setNames(c("accession", model)) %>%
      column_to_rownames("accession")
    fits_frame <- cbind(fits_frame, new_models[model])
  }
  return(fits_frame)
}

#' @name CALC.ModelGenerator_AllAccession_R2
#' @description Creates a list of adj. R^2 values over the entire dataset,
not just specific accessions.
#' @param data_df A dataframe containing all measured data points with IDs
and their measures.
#' @param fill_models_df A vector of model strings.
#' @returns A named list of adj. R^2 values as doubles.

CALC.ModelGenerator_AllAccessions_R2 <- function(data_df, fill_models_df) {
  list_out <- list()
  for(model in fill_models_df$models) {
    list_out <- list_out %>%
      append(summary(lm(model, data = data_df))[["adj.r.squared"]])
  }
  names(list_out) <- fill_models_df$models
  return(list_out)
}

#' @name CALC.ModelGenerator_AllAccessions_SBC
#' @description Creates a list of SBC values over the entire dataset, not
just specific accessions.
#' @param data_df A dataframe containing all measured data points with IDs
and their measures.
#' @param fill_models_df A vector of model strings.
#' @example A named list of SBC values as doubles.

CALC.ModelGenerator_AllAccessions_SBC <- function(data_df, fill_models_df) {
  list_out <- list()
  for(model in fill_models_df$models) {
    list_out <- list_out %>%
      append(BIC(lm(model, data = data_df)))
  }
  names(list_out) <- fill_models_df$models
  return(list_out)
```

```
}


# Tukey Test Family ----------------------------------------------------------

#' @name CALC.tukeyTester
#' @description Returns Tukey Test data on confidence intervals between
accessions for a single measure.
#' @param data_df A source datafile containing cactus pad IDs and their
corresponding measures.
#' @param measure The specific measure to test over all accessions.
#' @returns Tukey confidence interval dataframe of higher and lower interval
endpoints for each comparison requested.
#' names = factor, diff, lwr, upr, p adj
#' @example CALC.tukeyTester(SWITCHBOARD.csvMAINFILE, "height")

CALC.tukeyTester <- function(data_df, measure) {

  model=lm(data_df[[measure]] ~ factor(data_df$accession))
  ANOVA=aov(model)
  tukey_results <- TukeyHSD(x=ANOVA, 'factor(data_df$accession)',
conf.level=0.95)
  return(tukey_results)
}

#' @name CALC.tukeyLabel_DF
#' @description Labels accessions by groups that aren't significantly
different (using data from CALC.tukeyTester)
#' @param tukey_results A list of intervals generated by CALC.tukeyTester.
#' @returns A dataframe with two columns: a Letters column containing the
letters designating the groups that the accessions belong to, and the
accession column.

CALC.tukeyLabel_DF <- function(tukey_results) {
  Tukey.levels <- tukey_results[["factor(data_df$accession)"]][,4]
  Tukey.labels <- data.frame(multcompLetters(Tukey.levels)['Letters'])
  Tukey.labels$accession=rownames(Tukey.labels)
  rownames(Tukey.labels) <- NULL
  Tukey.labels=Tukey.labels[order(Tukey.labels$accession), ]
  return(Tukey.labels)
}

#' @name CALC.tukeyRelation
#' @description Breaks down a group relation dataframe generated by
CALC.tukeyTester into individual hypernode membership datapoints.
#' @param tukey_labels A two-column dataframe generated by CALC.tukeyTester
#' @returns A two-column dataframe - the first contains a single letter
(group designation), and the second column contains a single accession.

CALC.tukeyRelation <- function(tukey_labels) {
  tukey_relation <- tukey_labels
  tukey_relation$Letters <- sapply(tukey_relation$Letters, strsplit, split =
"")
  tukey_relation <- separate_rows(tukey_relation, Letters)
  tukey_relation <- tukey_relation[which(tukey_relation$Letters != ""),]
  return(tukey_relation)
}

#' @name CALC.tukeyEdgelist
#' @description Splits a edgelist dataframe into a vector of groups and the
accessions they contain, in a format that can be used to generate a
hypergraph.
#' @param tukey_relation A relation dataframe returned by CALC.tukeyRelation
```

```
#' @returns A named vector of groups (as letters) with collections of member
accessions as values.

CALC.tukeyEdgelist <- function(tukey_relation) {
  tukey_edgelist <- split(tukey_relation$accession, tukey_relation$Letters)
  return(tukey_edgelist)
}


# Model Acquisition and Filtration --------------------------------------

#' @name CALC.ModelFetch
#' @description
#' Retrieves a model from a dataframe constructed by a ModelGenerator() call,
using the accession and model_name passed.
#'   Will throw an error to the console if the requested model is not present
in the dataset.
#'   Will throw an error if the requested accession is not present in the
dataset.
#'
#' @param fits_df A dataframe created by a ModelGenerator() call, containing
an accession column and one or more model columns containing fitted models.
#' @param accession An integer corresponding to an accession from the
database.
#' @param model_name A string containing the model of interest to be
retrieved.
#' @example CALC.ModelFetch(DATA_OBJECTS.LinearRegressions_DF, 242,
"fresh_weight~width")

CALC.ModelFetch <- function(fits_df, accession, model_name) {
  if (!(model_name %in% colnames(fits_df)[-1])) {
    cat(paste0("The model '", model_name, "' is not present in ",
deparse(substitute(fits_df)), "\n"))
    cat(paste0("Available models include: ", colnames(fits_df)[-1], "\n"))
    return(NULL)
  }
  if (!(accession %in% fits_df$accession)) {
    cat("Invalid Accession Number: ", accession, "\n")
    return(NULL)
  }
  return(fits_df[[model_name]][which(SWITCHBOARD.ACCESSIONLIST ==
accession)][[1]])
}

#' @name CALC.ModelNumHighLeverage
#' @description Given a linear model, returns the number data points whose
leverage value exceeds a 3p/n threshold,
#'   where p is the number of parameters in the model,
#'   and n is the number of data points in the model.
#' @param model_in A linear model object.

CALC.ModelNumHighLeverage <- function(model_in) {
  threshold = 3*length(model_in$coefficients)/length(model_in$model[[1]])
  return(length(which(hatvalues(model_in) > threshold)))
}

#' @name CALC.ModelIndexHighLeverage
#' @description Given a linear model, returns a vector of indices of data
points whose leverage value exceeds a 3p/n threshold,
#'   where p is the number of parameters in the model,
#'   and n is the number of data points in the model.
#' @param model_in A linear model object.

CALC.ModelIndexHighLeverage <- function(model_in) {
```

```
   threshold <- 3*length(model_in$coefficients)/length(model_in$model[[1]])
   result_index <- as.vector(which(hatvalues(model_in) > threshold))
   return(result_index)
}

#' @name CALC.ModelValuePairsFromIndex
#' @description Given an index and a linear model object, retrieves the
values of the independent and dependent variable
#' @param result_index An integer corresponding to a datapoint in the model
object.
#' @param model_in A linear model object with a single predictor variable and
a single dependent variable.

CALC.ModelValuePairsFromIndex <- function(result_index, model_in) {
   if (length(result_index) == 0) {
     return(NULL)
   }
   model_data <- model_in$model
   cols <- names(model_data)
   return_values_df <- data.frame(matrix(ncol = 2, nrow =
length(result_index)))
   colnames(return_values_df) <- cols

   for (loc in length(result_index)) {
     pass_1 <- model_data[result_index[loc],][cols[1]]
     pass_2 <- model_data[result_index[loc],][cols[2]]
     row <- c(pass_1, pass_2)
     return_values_df[loc,] <- row
   }

   if (nrow(return_values_df) != 0) {
     return(return_values_df)
   } else {
     return(NULL)
   }
}

#' @name CALC.PadIDfromModelValuePairs
#' @description Given a dataframe of corresponding value pairs, a source
dataframe, and an accession number, assigns
#' @param value_pairs_df Value pair columns extracted from a linear model
object.
#' @param data_df A source datafile containing cactus pad IDs and their
corresponding measures.
#' @param accession An integer corresponding to an accession from the
database.

CALC.PadIDfromModelValuePairs <- function(value_pairs_df, data_df, accession)
{
   #print(length(nrow(value_pairs_df)))
   #print(value_pairs_df)
   if(nrow(value_pairs_df) == 0) {
     return(NULL)
   }
   pad_id_vec <- c()
   field_1 <- names(value_pairs_df)[1]
   field_2 <- names(value_pairs_df)[2]
   for(row in c(1:nrow(value_pairs_df))) {
     add_id <- data_df[
       which(
         data_df[[field_1]] == value_pairs_df[[field_1]][row] &
           data_df[[field_2]] == value_pairs_df[[field_2]][row] &
           data_df$accession == accession),
       ]$completeID
```

```
      pad_id_vec <- c(pad_id_vec, add_id)
  }
  #print(pad_id_vec)
  return(pad_id_vec)
}


# Summary Statistic Calculation -------------------------------------------

#' @name CALC.df_SBC
#' @description Generates a dataframe of Schwarz-Bayes Criterion values from
a dataframe of linear models. Each SBC value corresponds to the model in the
position it occupied.
#' @param fits_df A dataframe created by a ModelGenerator() call, containing
an accession column and one or more model columns containing fitted models.
#' @example CALC.df_SBC(DATA_OBJECTS.LinearRegressions_DF)

CALC.df_SBC <- function(fits_df) {
  return_df <- fits_df
  for (col in 2:ncol(fits_df)) {
    for (row in 1:nrow(fits_df)) {
      return_df[row, col] <- BIC(return_df[row, col][[1]])[[1]]
      #print(class(return_df[row, col]))
      #cat(ifelse(class(return_df[row, col]) == "list", "!", "&"))
    }
    return_df[,col] <- unlist(return_df[,col])
  }
  return(return_df)
}

#' @name CALC.df_R2
#' @description Generates a dataframe of Adjusted R^2 values from a dataframe
of linear models. Each Adj. R^2 value corresponds to the model in the
position it occupied.
#' @param fits_df A dataframe created by a ModelGenerator() call, containing
an accession column and one or more model columns containing fitted models.
#' @example CALC.df_R2(DATA_OBJECTS.LinearRegressions_DF)

CALC.df_R2 <- function(fits_df) {
  return_df <- fits_df
  for (col in 2:ncol(fits_df)) {
    for (row in 1:nrow(fits_df)) {
      return_df[row, col] <- summary(return_df[row,
col][[1]])[["adj.r.squared"]]
      #print(class(return_df[row, col]))
      #cat(ifelse(class(return_df[row, col]) == "list", "#", "@"))
    }
    return_df[,col] <- unlist(return_df[,col])
  }
  return(return_df)
}


# Result Comparison Functions ---------------------------------------------
# These functions do things slightly differently from previously defined
functions, so they take previously defined environment variables and do not
take any parameters.


#' @name CALC.CompareREIS_FRESHWEIGHT
#' @description Compares Reis et. al.'s regression for Cladode Fresh weight
on their data with the same formula type on the current data,
#' as well as the full multiplicative and elliptical models for fresh weight
created in this analysis.
```

```
CALC.CompareREIS_FRESHWEIGHT <- function() {
  DATA_OBJECTS.Comparison.REIS.freshweight <- data.frame(matrix(0.91, ncol =
1, nrow = 15)) %>%
    mutate(a =
DATA_OBJECTS.comparison_mult_R2_FILT$`fresh_weight~width*height*thickness`) %
>%
    mutate(b =
DATA_OBJECTS.comparison_mult_R2_FILT$`fresh_weight~width*height*diameter*thic
kness`) %>%
    mutate(c =
DATA_OBJECTS.comparison_ellip_R2_FILT$`fresh_weight~Theo_Area*thickness*Pade_
Derived_Diam`) %>%
    `rownames<-`(c(SWITCHBOARD.ACCESSIONLIST, "All")) %>%
    `colnames<-`(c("Reis L*W*T", "L*W*T", "L*W*T*D", "Full Ellip.")) %>%
    rownames_to_column("accession") %>%
    select(c(1, 3, 4, 5, 2)) %>%
    `rownames<-`(c(SWITCHBOARD.ACCESSIONLIST, "All"))

  UTIL.quickPNG("Reis-et-al FW Compare", "Heatmaps\\Graphs")

  heatmap_plot <- DATA_OBJECTS.Comparison.REIS.freshweight %>%
pivot_longer(cols = -1, names_to = "models", values_to = "values") %>%
    mutate(models = factor(models, levels =
colnames(DATA_OBJECTS.Comparison.REIS.freshweight))) %>%
    ggplot(aes(as.factor(models), as.factor(accession), fill = values)) +
    geom_tile() +
    xlab("Models") +
    ylab("Accession") +
    ggtitle("Reis-et-al FW Compare") +
    theme(axis.text.x = element_text(angle = 60, vjust = 0, hjust=0)) +
    scale_fill_viridis(direction = -1, begin =
max(DATA_OBJECTS.Comparison.REIS.freshweight[-1]), end =
ifelse(min(DATA_OBJECTS.Comparison.REIS.freshweight[-1]) < 0, 0,
min(DATA_OBJECTS.Comparison.REIS.freshweight[-1])))
  print(heatmap_plot)

  dev.off()

  UTIL.quickCSV(DATA_OBJECTS.Comparison.REIS.freshweight, "Reis-et-al FW
Compare", "Heatmaps\\Tables")
}

#' @name CALC.CompareREIS_AREA
#' @description Compares Reis et. al.'s regression for cladode area on their
data with the same formula type on the current data,
#' as well as the elliptical approximation and diameter-augmented
approximation (A_th*D) on the current data.

CALC.CompareREIS_AREA <- function() {

  DATA_OBJECTS.Comparison.REIS.area <-
DATA_OBJECTS.ComparativeRegressions_REIS_AREA_UNFILT %>%
    CALC.df_R2() %>%

UTIL.AllAccessionToMain(CALC.ModelGenerator_AllAccessions_R2(SWITCHBOARD.csvA
UGMFILE, DATA_OBJECTS.Comparison.REIS.area)) %>%
    `colnames<-`(c("accession",
SWITCHBOARD.models_to_compare.REIS.AREA_abbr)) %>%
    mutate("Reis Area" = 0.91)

  UTIL.quickPNG("Reis-et-al Area Compare", "Heatmaps\\Graphs")
```

```r
  heatmap_plot <- DATA_OBJECTS.Comparison.REIS.area %>% pivot_longer(cols = -
1, names_to = "models", values_to = "values") %>%
    mutate(models = factor(models, levels =
colnames(DATA_OBJECTS.Comparison.REIS.area))) %>%
    ggplot(aes(as.factor(models), as.factor(accession), fill = values)) +
    geom_tile() +
    xlab("Models") +
    ylab("Accession") +
    ggtitle("Reis-et-al Area Compare") +
    theme(axis.text.x = element_text(angle = 60, vjust = 0, hjust=0)) +
    scale_fill_viridis(direction = -1, begin =
max(DATA_OBJECTS.Comparison.REIS.area[-1]), end =
ifelse(min(DATA_OBJECTS.Comparison.REIS.area[-1]) < 0, 0,
min(DATA_OBJECTS.Comparison.REIS.area[-1])))
  print(heatmap_plot)

  dev.off()

  UTIL.quickCSV(DATA_OBJECTS.Comparison.REIS.area, "Reis-et-al Area Compare",
"Heatmaps\\Tables")
}

#' @name CALC.CompareREIS_DRYWEIGHT
#' @description
CALC.CompareREIS_DRYWEIGHT <- function() {
  DATA_OBJECTS.Comparison.REIS.dryweight <-
DATA_OBJECTS.ComparativeRegressions_REIS_DRYWEIGHT_UNFILT %>%
    CALC.df_R2() %>%

UTIL.AllAccessionToMain(CALC.ModelGenerator_AllAccessions_R2(SWITCHBOARD.csvA
UGMFILE, DATA_OBJECTS.Comparison.REIS.dryweight)) %>%
    `colnames<-`(c("accession",
SWITCHBOARD.models_to_compare.REIS.DRYWEIGHT_abbr)) %>%
    mutate("Reis DW" = 0.72)

  UTIL.quickPNG("Reis-et-al Dry Weight Compare", "Heatmaps\\Graphs")

  heatmap_plot <- DATA_OBJECTS.Comparison.REIS.dryweight %>%
pivot_longer(cols = -1, names_to = "models", values_to = "values") %>%
    mutate(models = factor(models, levels =
colnames(DATA_OBJECTS.Comparison.REIS.dryweight))) %>%
    ggplot(aes(as.factor(models), as.factor(accession), fill = values)) +
    geom_tile() +
    xlab("Models") +
    ylab("Accession") +
    ggtitle("Reis-et-al Dry Weight Compare") +
    theme(axis.text.x = element_text(angle = 60, vjust = 0, hjust=0)) +
    scale_fill_viridis(direction = -1, begin =
max(DATA_OBJECTS.Comparison.REIS.dryweight[-1]), end =
ifelse(min(DATA_OBJECTS.Comparison.REIS.dryweight[-1]) < 0, 0,
min(DATA_OBJECTS.Comparison.REIS.dryweight[-1])))
  print(heatmap_plot)

  dev.off()

  UTIL.quickCSV(DATA_OBJECTS.Comparison.REIS.dryweight, "Reis-et-al Dry
Weight Compare", "Heatmaps\\Tables")
}
```

A5.4 – GRAPHER_Functions.R

```
# qqGen and Intermeasure Graphing ----------------------------------------


#' @name GRAPHER.qqGen
#' @description Generates a Quantile-Quantile plot for the distribution of a
measure for a single accession.
#' @param data_df A dataframe containing all measured data points with IDs
and their measures.
#' @param accession A string containing a numerical accession ID.
#' @param measure A string containing the measure of interest.

GRAPHER.qqGen <- function(data_df, accession, measure) {
  target_data <- data_df[[measure]][data_df$accession == accession]
  UTIL.quickPNG(paste0(measure, "--", accession, "_LIN"), "QQPlot_Images")
  qqPlot(target_data, main = paste0(accession, ", ", measure), envelope =
0.95)
  dev.off()
}

#' @name GRAPHER.intermeasure
#' @description Generates a measure vs. measure plot of two measures for an
accession, with a fitted linear regression line.
#' @param fits_df A dataframe containing pregenerated models from
CALC.ModelGenerator.
#' @param accession A string containing a numerical accession ID.
#' @param modelname A string containing the regression model as text, in the
form "y~x"

GRAPHER.intermeasure <- function(fits_df, accession, modelname) {
  #GET x_measure and y_measure from model object
  model_in <- CALC.ModelFetch(fits_df, accession, modelname)

  model_in <- UTIL.AugmentModelWithLeverageTagging(model_in)

  model_in <- UTIL.AugmentModelWithID(model_in, accession)

  colnames_data <- names(model_in$model)
  model_df <- select(model_in$model, c(colnames_data))
  y_measure <- colnames_data[1]
  x_measure <- colnames_data[2]
  slope <- model_in$coefficients[2]
  intercept <- model_in$coefficients[1]

  adj_rsq <- round(summary(model_in)[["adj.r.squared"]], 3)

  intermeasure_plot_out <- ggplot(model_df,
aes(!!as.symbol(x_measure), !!as.symbol(y_measure),
                                            label = ifelse(tagged ==
"N/A", "", ID))) +
    geom_point(aes(color = tagged)) +
    xlab(x_measure) +
    ylab(y_measure) +
    labs(title = (paste0("INTERMEASURE\nMODEL:", modelname, "\nACCESSION:",
accession, "\nADJ.R^2:", adj_rsq))) +
    geom_text(hjust=1.05, vjust=0) +
    geom_text(hjust=-.20, vjust=0) +
    geom_abline(intercept = intercept, slope = slope)
  plot(intermeasure_plot_out)
}

#' @name GRAPHER.ViolinPlot
```

```
#' @description Generates a Violin plot of the range of valus of a measure
over all accessions.
#' @param data_df The table of measure values by individual pads.
#' @param measure The specific measure of interest in the data table.

GRAPHER.ViolinPlot <- function(data_df, measure) {
  violin_plot <- ggplot(data_df, aes(get(measure), as.factor(accession), fill
= as.factor(accession))) +
    geom_violin(trim=F, show.legend = F) +
    geom_boxplot(width = 0.2) +
    xlab(measure) +
    ylab("Accession") +
    ggtitle(paste0("Distribution of ", measure, " by Accession")) +
    theme(legend.position = "none")
  plot(violin_plot)
}

# Tukey Comparison and Hypergraphs ---------------------------------------

#' @name GRAPHER.tukeyPlotter
#' @description Generates a .png of a Tukey mean intercomparison interval
analysis between all possible pairs of accessions.
#' @param tukey_edgelist A group membership vector generated by
CALC.tukeyEdgelist

GRAPHER.tukeyPlotter <- function(tukey_edgelist) {
  tukeyplot <- plot(tukey_edgelist, las=1 , col="gray29")
  show(tukeyplot)
  #return(tukeyplot)
}

#' @name GRAPHER.tukeyHypergraph
#' @description Generates a Hypergraph where accessions with statistically
non-significant differences in their means share a hyperedge or group with
each other.
#' @param tukey_edgelist A group membership vector generated by
CALC.tukeyEdgelist
#' @param measure A string designating the measure to be analyzed - mostly to
aid in graph legend generating.

GRAPHER.tukeyHypergraph <- function(tukey_edgelist, measure_string) {
  plot(hypergraph_from_edgelist(tukey_edgelist))
  text(0, 1.25, cex = 1.75, paste0("Hypergraph of ", measure_string))
}


# Heatmaps of Summary Statistics ----------------------------------------

#' @name GRAPHER.R2heatmap
#' @description Generates a heatmap plot of the adj. R^2 values of all models
in a particular model set, fitted over all accessions.
#' @param R2_data_df A dataframe with accession rows and model columns
containing adj. R^2 values as doubles.
#' @param heatmap_title A string with the title of the heatmap.
#' @param abbreviations A vector of string abbreviations for the models for
intelligible graphing.

GRAPHER.R2heatmap <- function(R2_data_df, heatmap_title, abbreviations) {

  if (!(is.vector(abbreviations))) {
    abbreviations <- abbreviations$models
  }

  heatmap_plot <- R2_data_df %>%
```

```
    `colnames<-`(c("accession", abbreviations)) %>%
    pivot_longer(cols = -1, names_to = "models", values_to = "values") %>%
    mutate(models = factor(models, levels = abbreviations)) %>%
    ggplot(aes(as.factor(models), as.factor(accession), fill = values)) +
      geom_tile() +
      xlab("Models") +
      ylab("Accession") +
      ggtitle(heatmap_title) +
      theme(axis.text.x = element_text(angle = 60, vjust = 0, hjust=0)) +
      scale_fill_viridis(direction = -1, begin = max(R2_data_df[-1]), end =
ifelse(min(R2_data_df[-1]) < 0, 0, min(R2_data_df[-1])))
  print(heatmap_plot)
}


#' @name GRAPHER.SBCheatmap
#' @description Generates a heatmap plot of the SBC values of all models in a
particular model set, fitted over all accessions.
#' @param SBC_data_df A dataframe with accession rows and model columns
containing SBC values as doubles.
#' @param heatmap_title A string with the title of the heatmap.
#' @param abbreviations A vector of string abbreviations for the models for
intelligible graphing.

GRAPHER.SBCheatmap <- function(SBC_data_df, heatmap_title, abbreviations) {

  if (!(is.vector(abbreviations))) {
    abbreviations <- abbreviations$models
  }

  normal_table_by_row <- function(table_in) {
    for (row in c(1:nrow(table_in))) {
      vector_in <- table_in[row,]
      table_in[row,] <- (vector_in-min(vector_in))/max(vector_in)
      table_in[row,] <- table_in[row,] * 1/max(table_in[row,])
    }
    return(table_in)
  }


  heatmap_plot <- SBC_data_df %>%
    `rownames<-`(c(SWITCHBOARD.ACCESSIONLIST, "All")) %>%
    `colnames<-`(c("accession", abbreviations)) %>%
    {.[-1]} %>%
    normal_table_by_row() %>%
    rownames_to_column("accession") %>%
    `rownames<-`(rownames(SBC_data_df)) %>%
    pivot_longer(cols = -1, names_to = "models", values_to = "values") %>%
    mutate(models = factor(models, levels = abbreviations)) %>%
    ggplot(aes(models, as.ordered(accession), fill = values)) +
      geom_tile() +
      xlab("Models") +
      ylab("Accession") +
      ggtitle(heatmap_title) +
      theme(axis.text.x = element_text(angle = 60, vjust = 0, hjust=0)) +
      scale_fill_viridis(direction = -1, option = "A")
  print(heatmap_plot)
}




# 3D Scatterplots --------------------------------------------------------

GRAPHER.3DScatterPlot <- function(data_df, x_axis, y_axis, z_axis) {
```

```
plotpalette_3D <- c("#009a00", # Green
                    "#ef1101", # Bright Red
                    "#5c0300", # Burgundy
                    "#fd9206", # Pale Orange
                    "#a45d00", # Light Brown
                    "#fef636", # Pale Yellow
                    "#8f8a00", # Dirty Yellow
                    "#0006a4", # Dark Blue
                    "#0efd0e", # Bright Green
                    "#010afb", # Bright Blue
                    "#ba20fd", # Lavender
                    "#7600a8", # Purple
                    "#080f0f", # Black
                    "#faffff"  # White
                    ) #Fix these colors later

print(data_df["width"])

accession_vector <- data_df$accession
x_axis <- data_df[x_axis]
y_axis <- data_df[y_axis]
z_axis <- data_df[z_axis]

plot_ly(x = x_axis, y = y_axis, z = z_axis, color =
as.factor(accession_vector), colors =  plotpalette_3D)
}
```

A5.5 – COMPILE_Functions.R

```
# qqGen and Intermeasure Plots --------------------------------------------

#' @name COMPILE.qqGen
#' @description Generates and stores
#' all Quantile-Quantile plot for measures in a dataframe as .png files.
#' @param data_df The table of measure values by individual pads.
#' @param accessions A vector of accessions IDs.
#' @param measures A vector of measures to generate plots for.

COMPILE.qqGen <- function(data_df, accessions, measures) {

  for(accession in accessions) {
    for(measure in measures) {
      GRAPHER.qqGen(data_df, accession, measure)
    }
  }
}

#' @name COMPILE.intermeasure
#' @description Generates and stores all intermeasure plots for all measures
and accessions in a dataframe as .png files.
#' @param model_df The list of models to plot, by name.

COMPILE.intermeasure <- function(model_df) {

  modelnames = names(model_df)[-1]


  for(accession in SWITCHBOARD.ACCESSIONLIST) {
    cat(paste0("For Accession ", accession, "\n"))
    for (modelname in modelnames) {
      cat(paste0(modelname, "..."))
      UTIL.quickPNG(paste0(accession, "_", modelname, "_INTERMEASURE"),
"INTERMEASURE")
      GRAPHER.intermeasure(model_df, accession, modelname)
      dev.off()
    }
    cat("\n")
  }
}

# Tukey Mean Test Graphs --------------------------------------------------

#' @name COMPILE.tukeygraphs
#' @description Generates all mean intercomparison Tukey tests for a given
dataset, comparing between accessions.
#' @param data_df The table of measure values by individual pads.

COMPILE.tukeygraphs <- function(data_df) {
  measures <- c("width", "height", "diameter", "thickness", "fresh_weight")

  for (measure in measures) {
    UTIL.quickPNG(measure, "Tukey_Graphs\\Tukey_Intercomparisons")
    CALC.tukeyTester(data_df, measure) %T>%
      GRAPHER.tukeyPlotter() ->  tukey_results
    dev.off()

    UTIL.quickPNG(measure, "Tukey_Graphs\\Tukey_Hypergraphs")
    tukey_edgelist <- tukey_results %>%
      CALC.tukeyLabel_DF() %>%
      CALC.tukeyRelation() %>%
      CALC.tukeyEdgelist()
```

```
      GRAPHER.tukeyHypergraph(tukey_edgelist, measure)
      dev.off()
    }
}

# Summary Statistic Graphs -----------------------------------------------

#' @name COMPILE.R2heatmap
#' @description Generates and stores an adj. R^2 Heatmap for a dataframe of
values, by model and accession, as a .png file.
#' @param R2_df_in The dataframe of SBC values, with model columns and
accession rows.
#' @param filenamestring The name given for the .png file.
#' @param subfolder The folder that the file should be stored in.
#' @param abbreviations A vector of string abbreviations for the models for
intelligible graphing.

COMPILE.R2heatmap <- function(R2_df_in, filenamestring, subfolder,
heatmap_title, abbreviations) {
  UTIL.quickPNG(filenamestring, subfolder)
  GRAPHER.R2heatmap(R2_df_in, heatmap_title, abbreviations)
  dev.off()
}

#' @name COMPILE.SBCheatmap
#' @description Generates and stores a SBC Heatmap for a dataframe of values,
by model and accession, as a .png file.
#' @param SBC_df_in The dataframe of SBC values, with model columns and
accession rows.
#' @param filenamestring The name given for the .png file.
#' @param subfolder The folder that the file should be stored in.
#' @param abbreviations A vector of string abbreviations for the models for
intelligible graphing.

COMPILE.SBCheatmap <- function(SBC_df_in, filenamestring, subfolder,
heatmap_title, abbreviations) {
  UTIL.quickPNG(filenamestring, subfolder)
  GRAPHER.SBCheatmap(SBC_df_in, heatmap_title, abbreviations)
  dev.off()
}

#' @name COMPILE.ViolinPlot
#' @description Generates and stores a Violin plot of the range of values for
a particular set of measures across multiple accessions.
#' @param data_df The table of measure values by individual pads.
#' @param measurelist The list of measures to iterate over, as a character
vector.
#' @param data_title The title of the heatmap plot, as a string.

COMPILE.ViolinPlot <- function(data_df, measurelist, data_title) {
  for (measure in measurelist) {
    UTIL.quickPNG(paste0(measure, " - ", data_title),
"Violin_Graphs\\Graphs")
    GRAPHER.ViolinPlot(data_df, measure)
    dev.off()
  }
}
```

A5.6 – File Loading and Prep.R

```
#Load relevant packages into library

cat("Loading Packages...\n")
library("dplyr")
library("tibble")
library("rsq")
library("car")
library("Rcmdr")
library("hash")
library("multcompView")
library("HyperG")
library("plotly")
library("htmlwidgets")
library("magrittr")
library("tidyr")
library("multiApply")
library("stringr")
library("viridis")
library("pheatmap")
library("RColorBrewer")
library("data.table")


# SWITCHBOARD Setup ------------------------------------------------------

cat("Setting up SWITCHBOARD: \n")

#NOTE: ALL DATAFILES MUST BE LOADED INTO THE WORKING DIRECTORY PRIOR TO THE
PROGRAM START
SWITCHBOARD.DIRECTORY <- getwd()

cat("::| Loading Dataset of Interest...\n")
SWITCHBOARD.csvMAINFILE <- read.csv(file = DATA_FILE)

# This section hard-codes the dry_weight data collected in the experiment for
addition into the data set.

### USE THIS one as agument file
# https://community.rstudio.com/t/recognize-usr-bin-env-rscript-header-as-r-
script/59237
SWITCHBOARD.FRESH_DRY_RAWS <- tribble(
  ~accession, ~mean_freshweight, ~SEM_mean_fresh, ~mean_dryweight,
~SEM_mean_dry,
  242, 9.17, 0.94, 0.6, 0.07,
  246, 8.48, 0.94, 0.49, 0.07,
  319, 8.73, 0.94, 0.6, 0.07,
  325, 10.23, 1, 0.68, 0.07,
  326, 5.76, 0.94, 0.49, 0.07,
  390, 10.65, 0.97, 0.89, 0.07,
  572, 10.37, 0.94, 0.65, 0.07,
  580, 5.47, 0.91, 0.4, 0.07,
  582, 8.11, 0.94, 0.57, 0.07,
  584, 9.71, 1.13, 0.66, 0.08,
  585, 4.56, 0.95, 0.32, 0.07,
  839, 3.97, 0.94, 0.33, 0.07,
  845, 7.59, 0.94, 0.71, 0.07,
  854, 2.31, 0.94, 0.21, 0.07
)

cat("::| Setting Dry Weights...\n")
SWITCHBOARD.DRY_PROPORTIONS <- SWITCHBOARD.FRESH_DRY_RAWS %>%
  mutate(proportion = mean_dryweight/mean_freshweight) %>%
  select(accession, proportion)
```

```
cat("::| Setting Accession List...\n")
## This one could be read from the file load
SWITCHBOARD.ACCESSIONLIST <- c(242, 246, 319, 325, 326, 390, 572, 580, 582,
584, 585, 839, 845, 854)

## Save the CSV file
cat("::| Setting Derived Measures...\n")
SWITCHBOARD.csvAUGMFILE <- SWITCHBOARD.csvMAINFILE %>%
  mutate(H_div_W = height / width,
         FW_div_W = fresh_weight / width,
         FW_div_D = fresh_weight / diameter,
         FW_div_H = fresh_weight / height,
         FW_div_T = fresh_weight / thickness,
         D_div_W = diameter / width) %>%
  merge(read.csv('Pad_Area_Estimations.csv', fileEncoding = 'UTF-8-BOM'), by
= "completeID") %>%
  mutate(Theo_Area = pi*height*width*0.25) %>%
  mutate(PartRatio = ((height-width)/(height+width))^2) %>%
  mutate(Pade_Peri = pi*(height+width)*(64-3*PartRatio^2)/(64-
16*PartRatio)) %>%
  mutate(Pade_Derived_Diam = Pade_Peri/pi) %>%
  full_join(SWITCHBOARD.DRY_PROPORTIONS, by = "accession") %>%
  mutate(dry_weight = proportion*fresh_weight)

cat("::| Setting Model Guides...\n")
DATA_OBJECTS.MODELS_intermeasure <- tribble(
  ~ xname, ~ yname,
  'width',     'fresh_weight',
  'height',    'fresh_weight',
  'diameter',  'fresh_weight',
  'thickness', 'fresh_weight',
  'width',     'height',
  'width',     'diameter',
  'width',     'thickness',
  'height',    'diameter',
  'height',    'thickness',
  'diameter',  'thickness',
)

cat("::| Setting Intermeasure Models\n")
SWITCHBOARD.models_intermeasure_abbr <- c("FW~W", "FW~H", "FW~D", "FW~T",
"H~W", "D~W", "T~W", "D~H", "T~H", "T~D")

cat("::| Setting Initial Measures\n")
SWITCHBOARD.initialMeasures_VEC <- c("width", "diameter", "height",
"thickness", "fresh_weight")

cat("::| Setting Multiplicative Models for Rectangular Model Calculation\n")
SWITCHBOARD.models_multiplicative <- c(
  "fresh_weight~width", "fresh_weight~height",
"fresh_weight~diameter","fresh_weight~thickness",
  "fresh_weight~D_div_W",
  "fresh_weight~width+height", "fresh_weight~width*height",
  "fresh_weight~width+diameter", "fresh_weight~width*diameter",
  "fresh_weight~width+thickness", "fresh_weight~width*thickness",
  "fresh_weight~height+diameter", "fresh_weight~height*diameter",
  "fresh_weight~height+thickness", "fresh_weight~height*thickness",
  "fresh_weight~diameter+thickness", "fresh_weight~diameter*thickness",
  "fresh_weight~width+height+diameter", "fresh_weight~width*height*diameter",
  "fresh_weight~width+height+thickness",
"fresh_weight~width*height*thickness",
  "fresh_weight~width+diameter+thickness",
"fresh_weight~width*diameter*thickness",
```

```
    "fresh_weight~height+diameter+thickness",
"fresh_weight~height*diameter*thickness",
    "fresh_weight~width+height+diameter+thickness",
"fresh_weight~width*height*diameter*thickness"
)

cat("::| Setting Multiplicative Models for Elliptical Model Calculation\n")
SWITCHBOARD.models_multiplicative_RATIO <- c(
  "fresh_weight~width", "fresh_weight~height",
"fresh_weight~diameter","fresh_weight~thickness",
  "fresh_weight~D_div_W",
  "fresh_weight~width+height", "fresh_weight~width*height",
  "fresh_weight~width+D_div_W", "fresh_weight~width*D_div_W",
  "fresh_weight~width+thickness", "fresh_weight~width*thickness",
  "fresh_weight~height+D_div_W", "fresh_weight~height*D_div_W",
  "fresh_weight~height+thickness", "fresh_weight~height*thickness",
  "fresh_weight~D_div_W+thickness", "fresh_weight~D_div_W*thickness",
  "fresh_weight~width+height+D_div_W", "fresh_weight~width*height*D_div_W",
  "fresh_weight~width+height+thickness",
"fresh_weight~width*height*thickness",
  "fresh_weight~width+D_div_W+thickness",
"fresh_weight~width*D_div_W*thickness",
  "fresh_weight~height+D_div_W+thickness",
"fresh_weight~height*D_div_W*thickness",
  "fresh_weight~width+height+D_div_W+thickness",
"fresh_weight~width*height*D_div_W*thickness"
)

cat("::| Setting Multiplicative Abbreviations\n")
SWITCHBOARD.models_multiplicative_abbr <- c(
  "W", "H", "D", "T",
  "D/W",
  "W+H", "W*H",
  "W+D", "W*D",
  "W+T", "W*T",
  "H+D", "H*D",
  "H+T", "H*T",
  "D+T", "D*T",
  "W+H+D", "W*H*D",
  "W+H+T", "W*H*T",
  "W+D+T", "W*D*T",
  "H+D+T", "H*D*T",
  "W+H+D+T", "W*H*D*T"
  #"I(width*height*diameter*thickness)"
)

SWITCHBOARD.models_multiplicative_abbr_RATIO <- c(
  "W", "H", "D", "T",
  "D/W (r)",
  "W+H", "W*H",
  "W+r", "W*r",
  "W+T", "W*T",
  "H+r", "H*r",
  "H+T", "H*T",
  "T+r", "T*r",
  "W+H+r", "W*H*r",
  "W+H+T", "W*H*T",
  "W+T+r", "W*T*r",
  "H+T+r", "H*T*r",
  "W+H+T+r", "W*H*T*r"
)

cat("::| Setting Elliptical Abbreviations\n")
SWITCHBOARD.models_elliptical <- c(
```

```
    "fresh_weight~Area",
    "fresh_weight~Theo_Area",
    "fresh_weight~Area+thickness",
    "fresh_weight~Area*thickness",
    "fresh_weight~Theo_Area+thickness",
    "fresh_weight~Theo_Area*thickness",
    "fresh_weight~Area+thickness+diameter",
    "fresh_weight~Area*thickness*diameter",
    "fresh_weight~Theo_Area+thickness+Pade_Derived_Diam",
    "fresh_weight~Theo_Area*thickness*Pade_Derived_Diam"
)

SWITCHBOARD.models_elliptical_abbr <- c(
    "A", "Th_A",
    "A+T", "A*T",
    "Th_A+T", "Th_A*T",
    "A+T+D", "A*T*D",
    "Th_A+T+PD", "Th_A*T*PD"
)

#---
cat("::| Setting Reis et. al. Comparison Models\n")
SWITCHBOARD.models_to_compare.REIS.AREA <- c(
    "Area~height*width",
    "Area~Theo_Area",
    "Area~Theo_Area*diameter"
)

SWITCHBOARD.models_to_compare.REIS.AREA_abbr <- c(
    "A~HW",
    "A~A_th",
    "A~A_th*D"
)

SWITCHBOARD.models_to_compare.REIS.DRYWEIGHT <- c(
    "dry_weight~width*thickness",
    "dry_weight~width*thickness*Pade_Derived_Diam",
    "dry_weight~Theo_Area*thickness*Pade_Derived_Diam"
)

SWITCHBOARD.models_to_compare.REIS.DRYWEIGHT_abbr <- c(
    "DW~W*T",
    "DW~W*T*pdD",
    "DW~A_th*T*pdD"
)

#SWITCHBOARD.models_to_compare.REIS.DRY_WEIGHT

# DATA_OBJECT Generation: Unfiltered ---------------------------------------
----------
### https://stackoverflow.com/questions/16714020/loop-through-data-frame-and-
variable-names


cat("::| Polishing Model Guides...\n")
DATA_OBJECTS.MODELS_intermeasure$models <-
apply(DATA_OBJECTS.MODELS_intermeasure[, c("yname", "xname")], 1, paste,
collapse = "~")

DATA_OBJECTS.MODELS_multiplicative <-
as.data.frame(SWITCHBOARD.models_multiplicative) %>%
    set_colnames(c("models"))
```

```
DATA_OBJECTS.MODELS_multiplicative_abbr <-
as.data.frame(SWITCHBOARD.models_multiplicative_abbr) %>%
  set_colnames(c("models"))

DATA_OBJECTS.MODELS_multiplicative_RATIO <-
as.data.frame(SWITCHBOARD.models_multiplicative_RATIO) %>%
  set_colnames(c("models"))

DATA_OBJECTS.MODELS_multiplicative_abbr_RATIO <-
as.data.frame(SWITCHBOARD.models_multiplicative_abbr_RATIO) %>%
  set_colnames(c("models"))

DATA_OBJECTS.MODELS_elliptical <-
as.data.frame(SWITCHBOARD.models_elliptical) %>%
  set_colnames(c("models"))

DATA_OBJECTS.MODELS_elliptical_abbr <-
as.data.frame(SWITCHBOARD.models_elliptical_abbr) %>%
  set_colnames(c("models"))

DATA_OBJECTS.MODELS_REIS_AREA <-
as.data.frame(SWITCHBOARD.models_to_compare.REIS.AREA) %>%
  set_colnames(c("models"))

DATA_OBJECTS.MODELS_REIS_AREA_abbr <-
as.data.frame(SWITCHBOARD.models_to_compare.REIS.AREA_abbr) %>%
  set_colnames(c("models"))

DATA_OBJECTS.MODELS_REIS_DRYWEIGHT <-
as.data.frame(SWITCHBOARD.models_to_compare.REIS.DRYWEIGHT) %>%
  set_colnames(c("models"))

DATA_OBJECTS.MODELS_REIS_DRYWEIGHT_abbr <-
as.data.frame(SWITCHBOARD.models_to_compare.REIS.DRYWEIGHT_abbr) %>%
  set_colnames(c("models"))

DATA_OBJECTS.models_multiplicative.splitinteraction <-
UTIL.SplitModelListOnInteractions(DATA_OBJECTS.MODELS_multiplicative, 4)
DATA_OBJECTS.models_multiplicative.splitinteraction_RATIO <-
UTIL.SplitModelListOnInteractions(DATA_OBJECTS.MODELS_multiplicative_RATIO,
4)
DATA_OBJECTS.models_elliptical.splitinteraction <-
UTIL.SplitModelListOnInteractions(DATA_OBJECTS.MODELS_elliptical, 2)

DATA_OBJECTS.models_multiplicative_abbr.splitinteraction <-
UTIL.SplitModelListOnInteractions(DATA_OBJECTS.MODELS_multiplicative_abbr, 4)
DATA_OBJECTS.models_multiplicative_abbr.splitinteraction_RATIO <-
UTIL.SplitModelListOnInteractions(DATA_OBJECTS.MODELS_multiplicative_abbr_RAT
IO, 4)
DATA_OBJECTS.models_elliptical_abbr.splitinteraction <-
UTIL.SplitModelListOnInteractions(DATA_OBJECTS.MODELS_elliptical_abbr, 2)

cat("Generating Model Dataframes\n")

### loop
DATA_OBJECTS.LinearRegressions_DF <-
CALC.ModelGenerator(SWITCHBOARD.csvAUGMFILE,
DATA_OBJECTS.MODELS_intermeasure)
DATA_OBJECTS.MultiplicativeRegressions_DF_UNFILT <-
CALC.ModelGenerator(SWITCHBOARD.csvAUGMFILE,
DATA_OBJECTS.MODELS_multiplicative)
DATA_OBJECTS.MultiplicativeRegressions_DF_UNFILT_RATIO <-
CALC.ModelGenerator(SWITCHBOARD.csvAUGMFILE,
DATA_OBJECTS.MODELS_multiplicative_RATIO)
```

```
DATA_OBJECTS.EllipticalRegressions_DF_UNFILT <-
CALC.ModelGenerator(SWITCHBOARD.csvAUGMFILE, DATA_OBJECTS.MODELS_elliptical)

DATA_OBJECTS.ComparativeRegressions_REIS_AREA_UNFILT <-
CALC.ModelGenerator(SWITCHBOARD.csvAUGMFILE, DATA_OBJECTS.MODELS_REIS_AREA)

DATA_OBJECTS.ComparativeRegressions_REIS_DRYWEIGHT_UNFILT <-
CALC.ModelGenerator(SWITCHBOARD.csvAUGMFILE,
DATA_OBJECTS.MODELS_REIS_DRYWEIGHT)

# DATA_OBJECT Generation: Filtered -----------------------------------------
---
cat("Generating Filtered Datasets on 3p/n Threshold Value by Model\n")

cat("::| Writing Blacklists...\n")
DATA_OBJECTS.blacklist_3PN <-
UTIL.Assemble3PNDroplist(DATA_OBJECTS.LinearRegressions_DF[,-
which(names(DATA_OBJECTS.LinearRegressions_DF) %in% c("thickness~width",
"thickness~height", "thickness~diameter"))], SWITCHBOARD.csvMAINFILE)


cat("::| Writing Filtered Datasets...\n")
DATA_OBJECTS.csvAUGM_3PN <- UTIL.DropOnNPN(SWITCHBOARD.csvAUGMFILE,
DATA_OBJECTS.blacklist_3PN)

DATA_OBJECTS.ComparativeRegressions_REIS_AREA_3PN <-
CALC.ModelGenerator(DATA_OBJECTS.csvAUGM_3PN, DATA_OBJECTS.MODELS_REIS_AREA)

DATA_OBJECTS.ComparativeRegressions_REIS_DRYWEIGHT_3PN <-
CALC.ModelGenerator(DATA_OBJECTS.csvAUGM_3PN,
DATA_OBJECTS.MODELS_REIS_DRYWEIGHT)

# Summary Statistic Heatmap Generation -------------------------------------
----------------

cat("Generating Heatmaps and Writing CSV Data\n")
cat("::| Heatmaps...\n")
SWITCHBOARD.models_multiplicative.splitinteraction <-
UTIL.SplitModelListOnInteractions(DATA_OBJECTS.MODELS_multiplicative, 5)
SWITCHBOARD.models_multiplicative.splitinteraction_RATIO <-
UTIL.SplitModelListOnInteractions(DATA_OBJECTS.MODELS_multiplicative_RATIO,
5)
SWITCHBOARD.models_elliptical.splitinteraction <-
UTIL.SplitModelListOnInteractions(DATA_OBJECTS.MODELS_elliptical, 2)

SWITCHBOARD.models_multiplicative_abbr.splitinteraction <-
UTIL.SplitModelListOnInteractions(DATA_OBJECTS.MODELS_multiplicative_abbr, 5)
SWITCHBOARD.models_multiplicative_abbr.splitinteraction_RATIO <-
UTIL.SplitModelListOnInteractions(DATA_OBJECTS.MODELS_multiplicative_abbr_RAT
IO, 5)
SWITCHBOARD.models_elliptical_abbr.splitinteraction <-
UTIL.SplitModelListOnInteractions(DATA_OBJECTS.MODELS_elliptical_abbr, 2)

cat("::| CSVs...\n")
DATA_OBJECTS.comparison_mult_R2_FILT <- DATA_OBJECTS.csvAUGM_3PN %>%
  CALC.ModelGenerator(DATA_OBJECTS.MODELS_multiplicative) %>%
  CALC.df_R2() %>%

UTIL.AllAccessionToMain(CALC.ModelGenerator_AllAccessions_R2(DATA_OBJECTS.csv
AUGM_3PN, DATA_OBJECTS.MODELS_multiplicative))

DATA_OBJECTS.comparison_mult_R2_FILT_RATIO <- DATA_OBJECTS.csvAUGM_3PN %>%
  CALC.ModelGenerator(DATA_OBJECTS.MODELS_multiplicative_RATIO) %>%
  CALC.df_R2() %>%
```

```
UTIL.AllAccessionToMain(CALC.ModelGenerator_AllAccessions_R2(DATA_OBJECTS.csv
AUGM_3PN, DATA_OBJECTS.MODELS_multiplicative_RATIO))


DATA_OBJECTS.comparison_ellip_R2_FILT <- DATA_OBJECTS.csvAUGM_3PN %>%
  CALC.ModelGenerator(DATA_OBJECTS.MODELS_elliptical) %>%
  CALC.df_R2() %>%

UTIL.AllAccessionToMain(CALC.ModelGenerator_AllAccessions_R2(DATA_OBJECTS.csv
AUGM_3PN, DATA_OBJECTS.MODELS_elliptical))



DATA_OBJECTS.Comparison.REIS.area <-
as.data.frame(SWITCHBOARD.models_to_compare.REIS.AREA) %>%
  set_colnames(c("models"))

DATA_OBJECTS.Comparison.REIS.dryweight <-
as.data.frame(SWITCHBOARD.models_to_compare.REIS.DRYWEIGHT) %>%
  set_colnames(c("models"))
```

A5.7 – File Generation.R

```
cat("GENERATING PLOTS AND TABLES\n")

# Quantile-Quantile Plot Generation --------------------------------------

cat("Quantile-Quantile Plots\n")
COMPILE.qqGen(SWITCHBOARD.csvMAINFILE, SWITCHBOARD.ACCESSIONLIST,
SWITCHBOARD.initialMeasures_VEC)

# Intermeasure Analysis Figure Generation -------------------------------

cat("Intermeasure Plots\n")
COMPILE.intermeasure(DATA_OBJECTS.LinearRegressions_DF)

cat("Violin Plots\n")
COMPILE.ViolinPlot(SWITCHBOARD.csvMAINFILE, SWITCHBOARD.initialMeasures_VEC,
"Unfiltered")
COMPILE.ViolinPlot(DATA_OBJECTS.csvAUGM_3PN, SWITCHBOARD.initialMeasures_VEC,
"Filtered")

cat("Tukey Intercomparisons and Grouping Hypergraphs\n")
COMPILE.tukeygraphs(SWITCHBOARD.csvAUGMFILE)


# Heatmap Generation ----------------------------------------------------

cat("HEATMAP: Intermeasure Comparisons\n")
UTIL.StoreHeatmapsAndFile(SWITCHBOARD.csvAUGMFILE,
DATA_OBJECTS.MODELS_intermeasure, "Lin_UNF",   "Heatmaps", "Intermeasure
Models\n", SWITCHBOARD.models_intermeasure_abbr)


cat("HEATMAP: Multiplicative Model Comparisons - Unfiltered\n")
cat("With Interactions...")
UTIL.StoreHeatmapsAndFile(SWITCHBOARD.csvAUGMFILE,
DATA_OBJECTS.models_multiplicative.splitinteraction[[2]], "Mult_UNF_mult",
"Heatmaps", "Box Models\nUnfiltered Dataset with Interactions",
DATA_OBJECTS.models_multiplicative_abbr.splitinteraction[[2]]$models)
cat("Without Interactions...")
UTIL.StoreHeatmapsAndFile(SWITCHBOARD.csvAUGMFILE,
DATA_OBJECTS.models_multiplicative.splitinteraction[[1]], "Mult_UNF_plus",
"Heatmaps", "Box Models\nUnfiltered Dataset with No Interactions",
DATA_OBJECTS.models_multiplicative_abbr.splitinteraction[[1]]$models)
cat("Altogether...")
cat("\n")
UTIL.StoreHeatmapsAndFile(SWITCHBOARD.csvAUGMFILE,
DATA_OBJECTS.MODELS_multiplicative, "Mult_UNF",   "Heatmaps", "Box
Models\nUnfiltered Dataset",SWITCHBOARD.models_multiplicative_abbr)


cat("HEATMAP: Multiplicative Model Comparisons - FILTERED on 3P/N
Leverage\n")
cat("With Interactions...")
UTIL.StoreHeatmapsAndFile(DATA_OBJECTS.csvAUGM_3PN,
DATA_OBJECTS.models_multiplicative.splitinteraction[[2]], "Mult_3PN_mult",
"Heatmaps", "Box Models\n3P/N Dataset with Interactions",
DATA_OBJECTS.models_multiplicative_abbr.splitinteraction[[2]]$models)
cat("Without Interactions...")
UTIL.StoreHeatmapsAndFile(DATA_OBJECTS.csvAUGM_3PN,
DATA_OBJECTS.models_multiplicative.splitinteraction[[1]], "Mult_3PN_plus",
"Heatmaps", "Box Models\n3P/N Dataset with No Interactions",
DATA_OBJECTS.models_multiplicative_abbr.splitinteraction[[1]]$models)
cat("Altogether...")
```

```
cat("\n")
UTIL.StoreHeatmapsAndFile(DATA_OBJECTS.csvAUGM_3PN,
DATA_OBJECTS.MODELS_multiplicative, "Mult_3PN",  "Heatmaps", "Box
Models\n3P/N Dataset", SWITCHBOARD.models_multiplicative_abbr)

cat("HEATMAP: Multiplicative Model Comparisons - Unfiltered - Alternate
Formula\n")
cat("With Interactions...")
UTIL.StoreHeatmapsAndFile(SWITCHBOARD.csvAUGMFILE,
DATA_OBJECTS.models_multiplicative.splitinteraction_RATIO[[2]],
"Mult_UNF_mult_ALT",  "Heatmaps", "Box Models ALT\nUnfiltered Dataset with
Interactions",
DATA_OBJECTS.models_multiplicative_abbr.splitinteraction_RATIO[[2]]$models)
cat("Without Interactions...")
UTIL.StoreHeatmapsAndFile(SWITCHBOARD.csvAUGMFILE,
DATA_OBJECTS.models_multiplicative.splitinteraction_RATIO[[1]],
"Mult_UNF_plus_ALT",  "Heatmaps", "Box Models ALT\nUnfiltered Dataset with No
Interactions",
DATA_OBJECTS.models_multiplicative_abbr.splitinteraction_RATIO[[1]]$models)
cat("Altogether...")
cat("\n")
UTIL.StoreHeatmapsAndFile(SWITCHBOARD.csvAUGMFILE,
DATA_OBJECTS.MODELS_multiplicative_RATIO, "Mult_UNF_ALT",  "Heatmaps", "Box
Models ALT\nUnfiltered Dataset",SWITCHBOARD.models_multiplicative_abbr_RATIO)


cat("HEATMAP: Multiplicative Model Comparisons - FILTERED on 3P/N Leverage -
Alternate Formula\n")
cat("With Interactions...")
UTIL.StoreHeatmapsAndFile(DATA_OBJECTS.csvAUGM_3PN,
DATA_OBJECTS.models_multiplicative.splitinteraction_RATIO[[2]],
"Mult_3PN_mult_ALT",  "Heatmaps", "Box Model ALTs\n3P/N Dataset with
Interactions",
DATA_OBJECTS.models_multiplicative_abbr.splitinteraction_RATIO[[2]]$models)
cat("Without Interactions...")
UTIL.StoreHeatmapsAndFile(DATA_OBJECTS.csvAUGM_3PN,
DATA_OBJECTS.models_multiplicative.splitinteraction_RATIO[[1]],
"Mult_3PN_plus_ALT",  "Heatmaps", "Box Models ALT\n3P/N Dataset with No
Interactions",
DATA_OBJECTS.models_multiplicative_abbr.splitinteraction_RATIO[[1]]$models)
cat("Altogether...")
cat("\n")
UTIL.StoreHeatmapsAndFile(DATA_OBJECTS.csvAUGM_3PN,
DATA_OBJECTS.MODELS_multiplicative_RATIO, "Mult_3PN_ALT",  "Heatmaps", "Box
Models ALT\n3P/N Dataset", SWITCHBOARD.models_multiplicative_abbr_RATIO)


cat("HEATMAP: Elliptical Model Comparisons - Unfiltered\n")
cat("With Interactions...")
UTIL.StoreHeatmapsAndFile(SWITCHBOARD.csvAUGMFILE,
DATA_OBJECTS.models_elliptical.splitinteraction[[2]], "Ellip_UNF_mult",
"Heatmaps", "Fitting and Elliptical Models\nUnfiltered Dataset with
Interactions",
DATA_OBJECTS.models_elliptical_abbr.splitinteraction[[2]]$models)
cat("Without Interactions...")
UTIL.StoreHeatmapsAndFile(SWITCHBOARD.csvAUGMFILE,
DATA_OBJECTS.models_elliptical.splitinteraction[[1]], "Ellip_UNF_plus",
"Heatmaps", "Fitting and Elliptical Models\nUnfiltered Dataset with No
Interactions",
DATA_OBJECTS.models_elliptical_abbr.splitinteraction[[1]]$models)
cat("Altogether...")
cat("\n")
```

```
        UTIL.StoreHeatmapsAndFile(SWITCHBOARD.csvAUGMFILE,
        DATA_OBJECTS.MODELS_elliptical, "Ellip_UNF", "Heatmaps", "Fitting and
        Elliptical Models\nUnfiltered Dataset", SWITCHBOARD.models_elliptical_abbr)


        cat("HEATMAP: Elliptical Model Comparisons - FILTERED on 3P/N Leverage\n")
        cat("With Interactions...")
        UTIL.StoreHeatmapsAndFile(DATA_OBJECTS.csvAUGM_3PN,
        DATA_OBJECTS.models_elliptical.splitinteraction[[2]], "Ellip_3PN_mult",
        "Heatmaps", "Fitting and Elliptical Models\n3P/N Dataset with Interactions",
        DATA_OBJECTS.models_elliptical_abbr.splitinteraction[[2]]$models)
        cat("Without Interactions...")
          UTIL.StoreHeatmapsAndFile(DATA_OBJECTS.csvAUGM_3PN,
        DATA_OBJECTS.models_elliptical.splitinteraction[[1]], "Ellip_3PN_plus",
        "Heatmaps", "Fitting and Elliptical Models\n3P/N Dataset with No
        Interactions",
        DATA_OBJECTS.models_elliptical_abbr.splitinteraction[[1]]$models)
        cat("Altogether...")
        cat("\n")
        UTIL.StoreHeatmapsAndFile(DATA_OBJECTS.csvAUGM_3PN,
        DATA_OBJECTS.MODELS_elliptical, "Ellip_3PN", "Heatmaps", "Fitting and
        Elliptical Models\n3P/N Data", SWITCHBOARD.models_elliptical_abbr)


        cat("HEATMAP: Reis et. al. Result Comparisons\n")
        cat("Fresh Weight...")
        CALC.CompareREIS_FRESHWEIGHT()
        cat("Area...")
        CALC.CompareREIS_AREA()
        cat("Dry Weight...")
        cat("\n")
        CALC.CompareREIS_DRYWEIGHT()
        cat("\n")

        cat("Hard-Coded Heatmap for Power Law Investigation\n")

        powerlawtable <- data.frame(matrix(ncol=6,nrow=0, dimnames=list(NULL,
        c("dataset", "accession", "model_text", "abbreviation", "adjR2", "SBC"))))
        power_model_names <- c("log(height)", "log(width)", "log(height*width)")
        power_model_abbre <- c("H^n", "W^n", "(HW)^n")
        for (model_num in c(1:length(power_model_names)))
          {
          for (accession_num in c("All", SWITCHBOARD.ACCESSIONLIST)) {
            model_string <- paste0("log(Area)~",power_model_names[model_num])
            if (accession_num == "All") {
              #print("GIG")
              data_slice_UNF <- SWITCHBOARD.csvAUGMFILE
              data_slice_FIL <- DATA_OBJECTS.csvAUGM_3PN
            } else {
              data_slice_UNF <- SWITCHBOARD.csvAUGMFILE %>% filter(accession ==
        accession_num)
              data_slice_FIL <- DATA_OBJECTS.csvAUGM_3PN %>% filter(accession ==
        accession_num)
            }
            print(nrow(data_slice_UNF))
            model_object_UNF <- lm(model_string, data_slice_UNF)
            model_adjr_2_UNF <- as.numeric(summary(model_object_UNF)$adj.r.squared)
            model_SBCstt_UNF <- as.numeric(BIC(model_object_UNF))
            model_object_FIL <- lm(model_string, data_slice_FIL)
            model_adjr_2_FIL <- as.numeric(summary(model_object_FIL)$adj.r.squared)
            model_SBCstt_FIL <- as.numeric(BIC(model_object_FIL))
            powerlawtable[nrow(powerlawtable)+1,] <- c("Unfiltered",
        accession_num, model_string, power_model_abbre[model_num], model_adjr_2_UNF,
        model_SBCstt_UNF)
```

```
      powerlawtable[nrow(powerlawtable)+1,] <- c("Filtered 3P/N",
accession_num, model_string, power_model_abbre[model_num], model_adjr_2_FIL,
model_SBCstt_FIL)
   }
}
powerlawtable$adjR2 <- as.numeric(powerlawtable$adjR2)
powerlawtable$SBC <- as.numeric(powerlawtable$SBC)

# Adj. R^2 Power Law Heatmap Generation

UTIL.quickPNG("Power_Law_Models_Area_UNF_AdjR2", "Heatmaps\\Graphs")
placeholder_figure <- powerlawtable %>%
  filter(dataset == "Unfiltered") %>%
  select(accession, abbreviation, adjR2) %>%
  mutate(abbreviation = factor(abbreviation, levels = c("H^n", "W^n",
"(HW)^n"))) %>%
  ggplot(aes(abbreviation, as.factor(accession), fill = adjR2)) +
    geom_tile() +
    xlab("Models") +
    ylab("Accession") +
    ggtitle("Power Law Models - Area\nUnfiltered Data Adj. R^2") +
    theme(axis.text.x = element_text(angle = 60, vjust = 0, hjust=0)) +
    scale_fill_viridis(direction = 1)
print(placeholder_figure)
dev.off()
powerlawtable %>%
  filter(dataset == "Unfiltered") %>%
  select(accession, abbreviation, adjR2) %>%
  mutate(abbreviation = factor(abbreviation, levels = c("H^n", "W^n",
"(HW)^n"))) %>%
  pivot_wider(names_from = abbreviation, values_from = adjR2) %>%
  UTIL.quickCSV("Power_Law_Models_Area_UNF_AdjR2", "Heatmaps\\Tables")

UTIL.quickPNG("Power_Law_Models_Area_FILT_AdjR2", "Heatmaps\\Graphs")
placeholder_figure <- powerlawtable %>%
  filter(dataset == "Filtered 3P/N") %>%
  select(accession, abbreviation, adjR2) %>%
  mutate(abbreviation = factor(abbreviation, levels = c("H^n", "W^n",
"(HW)^n")))%>%
  ggplot(aes(abbreviation, as.factor(accession), fill = adjR2)) +
  geom_tile() +
  xlab("Models") +
  ylab("Accession") +
  ggtitle("Power Law Models - Area\nFiltered Data Adj. R^2") +
  theme(axis.text.x = element_text(angle = 60, vjust = 0, hjust=0)) +
  scale_fill_viridis(direction = 1)
print(placeholder_figure)
dev.off()
powerlawtable %>%
  filter(dataset == "Filtered 3P/N") %>%
  select(accession, abbreviation, adjR2) %>%
  mutate(abbreviation = factor(abbreviation, levels = c("H^n", "W^n",
"(HW)^n"))) %>%
  pivot_wider(names_from = abbreviation, values_from = adjR2) %>%
  UTIL.quickCSV("Power_Law_Models_Area_FILT_AdjR2", "Heatmaps\\Tables")

#SBC Power Law Heatmap Generation

UTIL.quickPNG("Power_Law_Models_Area_UNF_SBC", "Heatmaps\\Graphs")
placeholder_figure <- powerlawtable %>%
  filter(dataset == "Unfiltered") %>%
  select(accession, abbreviation, SBC) %>%
  mutate(abbreviation = factor(abbreviation, levels = c("H^n", "W^n",
"(HW)^n"))) %>%
```

```r
  group_by(accession) %>%
  mutate(SBC = SBC-min(SBC)) %>%
  mutate(SBC = SBC/max(SBC)) %>%
  ungroup() %>%
  ggplot(aes(abbreviation, as.factor(accession), fill = SBC)) +
  geom_tile() +
  xlab("Models") +
  ylab("Accession") +
  ggtitle("Power Law Models - Area\nUnfiltered Data Adj. SBC") +
  theme(axis.text.x = element_text(angle = 60, vjust = 0, hjust=0)) +
  scale_fill_viridis(direction = -1, option = "A")
print(placeholder_figure)
dev.off()
powerlawtable %>%
  filter(dataset == "Unfiltered") %>%
  select(accession, abbreviation, SBC) %>%
  mutate(abbreviation = factor(abbreviation, levels = c("H^n", "W^n",
"(HW)^n"))) %>%
  pivot_wider(names_from = abbreviation, values_from = SBC) %>%
  UTIL.quickCSV("Power_Law_Models_Area_UNF_SBC", "Heatmaps\\Tables")


UTIL.quickPNG("Power_Law_Models_Area_FILT_SBC", "Heatmaps\\Graphs")
placeholder_figure <- powerlawtable %>%
  filter(dataset == "Filtered 3P/N") %>%
  select(accession, abbreviation, SBC) %>%
  mutate(abbreviation = factor(abbreviation, levels = c("H^n", "W^n",
"(HW)^n"))) %>%
  group_by(accession) %>%
  mutate(SBC = SBC-min(SBC)) %>%
  mutate(SBC = SBC/max(SBC)) %>%
  ungroup() %>%
  ggplot(aes(abbreviation, as.factor(accession), fill = SBC)) +
  geom_tile() +
  xlab("Models") +
  ylab("Accession") +
  ggtitle("Power Law Models - Area\nFiltered Data SBC") +
  theme(axis.text.x = element_text(angle = 60, vjust = 0, hjust=0)) +
  scale_fill_viridis(direction = -1, option = "A")
print(placeholder_figure)
dev.off()
powerlawtable %>%
  filter(dataset == "Filtered 3P/N") %>%
  select(accession, abbreviation, SBC) %>%
  mutate(abbreviation = factor(abbreviation, levels = c("H^n", "W^n",
"(HW)^n"))) %>%
  pivot_wider(names_from = abbreviation, values_from = SBC) %>%
  UTIL.quickCSV("Power_Law_Models_Area_FILT_SBC", "Heatmaps\\Tables")

#-------------------------------------


gammalawtable <- data.frame(matrix(ncol=6,nrow=0, dimnames=list(NULL,
c("dataset", "accession", "model_text", "abbreviation", "adjR2", "SBC"))))
gamma_model_names <- c("height", "width", "thickness", "I(height*width)",
"I(height*width)+thickness")
gamma_model_abbre <- c("H", "W", "T", "HW", "HW+T")
for (model_num in c(1:length(gamma_model_names)))
{
  for (accession_num in c("All", SWITCHBOARD.ACCESSIONLIST)) {
    model_string <- paste0("fresh_weight~",gamma_model_names[model_num])
    if (accession_num == "All") {
      data_slice_UNF <- SWITCHBOARD.csvAUGMFILE
      data_slice_FIL <- DATA_OBJECTS.csvAUGM_3PN
```

```
    } else {
        data_slice_UNF <- SWITCHBOARD.csvAUGMFILE %>% filter(accession ==
accession_num)
        data_slice_FIL <- DATA_OBJECTS.csvAUGM_3PN %>% filter(accession ==
accession_num)
    }
    print(nrow(data_slice_UNF))
    model_object_UNF <- lm(model_string, data_slice_UNF)
    model_adjr_2_UNF <- as.numeric(summary(model_object_UNF)$adj.r.squared)
    model_SBCstt_UNF <- as.numeric(BIC(model_object_UNF))
    model_object_FIL <- lm(model_string, data_slice_FIL)
    model_adjr_2_FIL <- as.numeric(summary(model_object_FIL)$adj.r.squared)
    model_SBCstt_FIL <- as.numeric(BIC(model_object_FIL))
    gammalawtable[nrow(gammalawtable)+1,] <- c("Unfiltered",
accession_num, model_string, gamma_model_abbre[model_num], model_adjr_2_UNF,
model_SBCstt_UNF)
    gammalawtable[nrow(gammalawtable)+1,] <- c("Filtered 3P/N",
accession_num, model_string, gamma_model_abbre[model_num], model_adjr_2_FIL,
model_SBCstt_FIL)
    }
}
gammalawtable$adjR2 <- as.numeric(gammalawtable$adjR2)
gammalawtable$SBC <- as.numeric(gammalawtable$SBC)

#----------------------------

UTIL.quickPNG("Gamma_Law_Models_Fresh-Weight_UNF_AdjR2", "Heatmaps\\Graphs")
placeholder_figure <- gammalawtable %>%
    filter(dataset == "Unfiltered") %>%
    select(accession, abbreviation, adjR2) %>%
    mutate(abbreviation = factor(abbreviation, levels = c("H", "W", "T", "HW",
"HW+T"))) %>%
    ggplot(aes(abbreviation, as.factor(accession), fill = adjR2)) +
    geom_tile() +
    xlab("Models") +
    ylab("Accession") +
    ggtitle("Gamma Law Models - Fresh-Weight\nUnfiltered Data Adj. R^2") +
    theme(axis.text.x = element_text(angle = 60, vjust = 0, hjust=0)) +
    scale_fill_viridis(direction = 1)
print(placeholder_figure)
dev.off()
gammalawtable %>%
    filter(dataset == "Unfiltered") %>%
    select(accession, abbreviation, adjR2) %>%
    mutate(abbreviation = factor(abbreviation, levels = c("H", "W", "T", "HW",
"HW+T"))) %>%
    pivot_wider(names_from = abbreviation, values_from = adjR2) %>%
    UTIL.quickCSV("Gamma_Law_Models_Fresh-Weight_UNF_AdjR2",
"Heatmaps\\Tables")

UTIL.quickPNG("Gamma_Law_Models_Fresh-Weight_FILT_AdjR2", "Heatmaps\\Graphs")
placeholder_figure <- gammalawtable %>%
    filter(dataset == "Filtered 3P/N") %>%
    select(accession, abbreviation, adjR2) %>%
    mutate(abbreviation = factor(abbreviation, levels = c("H", "W", "T", "HW",
"HW+T")))%>%
    ggplot(aes(abbreviation, as.factor(accession), fill = adjR2)) +
    geom_tile() +
    xlab("Models") +
    ylab("Accession") +
    ggtitle("Gamma Law Models - Fresh-Weight\nFiltered Data Adj. R^2") +
    theme(axis.text.x = element_text(angle = 60, vjust = 0, hjust=0)) +
    scale_fill_viridis(direction = 1)
print(placeholder_figure)
```

```
dev.off()
gammalawtable %>%
  filter(dataset == "Filtered 3P/N") %>%
  select(accession, abbreviation, adjR2) %>%
  mutate(abbreviation = factor(abbreviation, levels = c("H", "W", "T", "HW",
"HW+T"))) %>%
  pivot_wider(names_from = abbreviation, values_from = adjR2) %>%
  UTIL.quickCSV("Gamma_Law_Models_Fresh-Weight_FILT_AdjR2",
"Heatmaps\\Tables")

#SBC Gamma Law Heatmap Generation

UTIL.quickPNG("Gamma_Law_Models_Fresh-Weight_UNF_SBC", "Heatmaps\\Graphs")
placeholder_figure <- gammalawtable %>%
  filter(dataset == "Unfiltered") %>%
  select(accession, abbreviation, SBC) %>%
  mutate(abbreviation = factor(abbreviation, levels = c("H", "W", "T", "HW",
"HW+T"))) %>%
  group_by(accession) %>%
  mutate(SBC = SBC-min(SBC)) %>%
  mutate(SBC = SBC/max(SBC)) %>%
  ungroup() %>%
  ggplot(aes(abbreviation, as.factor(accession), fill = SBC)) +
  geom_tile() +
  xlab("Models") +
  ylab("Accession") +
  ggtitle("Gamma Law Models - Fresh-Weight\nUnfiltered Data Adj. SBC") +
  theme(axis.text.x = element_text(angle = 60, vjust = 0, hjust=0)) +
  scale_fill_viridis(direction = -1, option = "A")
print(placeholder_figure)
dev.off()
gammalawtable %>%
  filter(dataset == "Unfiltered") %>%
  select(accession, abbreviation, SBC) %>%
  mutate(abbreviation = factor(abbreviation, levels = c("H", "W", "T", "HW",
"HW+T"))) %>%
  pivot_wider(names_from = abbreviation, values_from = SBC) %>%
  UTIL.quickCSV("Gamma_Law_Models_Fresh-Weight_UNF_SBC", "Heatmaps\\Tables")


UTIL.quickPNG("Gamma_Law_Models_Fresh-Weight_FILT_SBC", "Heatmaps\\Graphs")
placeholder_figure <- gammalawtable %>%
  filter(dataset == "Filtered 3P/N") %>%
  select(accession, abbreviation, SBC) %>%
  mutate(abbreviation = factor(abbreviation, levels = c("H", "W", "T", "HW",
"HW+T"))) %>%
  group_by(accession) %>%
  mutate(SBC = (SBC-min(SBC))) %>%
  mutate(SBC = SBC/max(SBC)) %>%
  ungroup() %>%
  ggplot(aes(abbreviation, as.factor(accession), fill = SBC)) +
  geom_tile() +
  xlab("Models") +
  ylab("Accession") +
  ggtitle("Gamma Law Models - Fresh-Weight\nFiltered Data SBC") +
  theme(axis.text.x = element_text(angle = 60, vjust = 0, hjust=0)) +
  scale_fill_viridis(direction = -1, option = "A")
print(placeholder_figure)
dev.off()
gammalawtable %>%
  filter(dataset == "Filtered 3P/N") %>%
  select(accession, abbreviation, SBC) %>%
  mutate(abbreviation = factor(abbreviation, levels = c("H", "W", "T", "HW",
"HW+T"))) %>%
```

```
    pivot_wider(names_from = abbreviation, values_from = SBC) %>%
    UTIL.quickCSV("Gamma_Law_Models_Fresh-Weight_FILT_SBC", "Heatmaps\\Tables")

# ggplot(aes(as.factor(models), as.factor(accession), fill = values)) +
#    geom_tile() +
# xlab("Models") +
# ylab("Accession") +
# ggtitle(heatmap_title) +
# theme(axis.text.x = element_text(angle = 60, vjust = 0, hjust=0)) +
    # scale_fill_viridis(direction = -1, begin = max(R2_data_df[-1]), end =
ifelse(min(R2_data_df[-1]) < 0, 0, min(R2_data_df[-1])))

cat("-------------------------------\n")
cat("-------------------------------\n")
```