Reflective essay
Daniel H. Chitwood (he/him)[1]

Perspective matters: a topological view of plant morphology
and training in the computational plant sciences[2]

> **Axiom 1.** Mathematical potential is equally present in different groups, irrespective of geographic, demographic, and economic boundaries.
>
> **Axiom 2.** Everyone can have joyful, meaningful, and empowering mathematical experiences.
>
> **Axiom 3.** Mathematics is a powerful, malleable tool that can be shaped and used differently by various communities to serve their needs.
>
> **Axiom 4.** Every student deserves to be treated with dignity and respect.
>
> Frederico Ardila Mantilla[3]

I am a plant developmental biologist who studies leaf shape, but as I will describe in this essay, needing to quantify the plant form, I began working with computational mathematicians, through a collaboration with Dr. Elizabeth Munch in Dept. Computational Mathematics, Science & Engineering (CMSE), applying Topological Data Analysis (TDA) techniques. The motto of TDA is that *shape is data, and data is shape*. Mathematical proofs from TDA state that the total information embedded within data structures can be extracted and summarized, with implications for how we measure phenotypic data, or any data, for that matter. I will describe geometric approaches I use to quantify leaves that led me to TDA to quantify more complex data structures, like X-ray Computed Tomography (CT) scans. To define topological features (like connected components, loops, or voids) TDA uses a lens function (for review see Amézquita et al., 2020[4]). A lens function is a real number value assigned to each data point in the dataset. A lens function can be anything. Far from being arbitrary, a lens function can be used as a mathematically defined perspective to view the structure of data from a chosen vantage point providing focus and insight into an analysis. Proofs state that to truly measure the entirety of information embedded in a data structure, that we need to analyze the shape of data from an infinite number of perspectives.

For me, at the heart of the relationship between plants and people is how we view them. The human mind, still, outperforms the most sophisticated algorithms and levels of computation, and is foundational in creating any type of training data, effortlessly taking in shapes, forms, and colors, distinguishing species and types of plants, and making observations. But each person—each culture—has a unique relationship with plants and views them differently. As expressed by TDA, to view the structure of data from only one perspective is incomplete. I opened this essay with the four axioms from Frederico Ardila

---

[2] For transparency into the tenure process for others, a copy of my reflective essay and teaching portfolio can be found at the following link:
[3] Ardila-Mantilla, Federico. "Todos cuentan: Cultivating diversity in combinatorics." *Notices of the AMS* 63.10 (2016): 1164-1170.
[4] References without a footnote indicate that I am an author on the publication and can be found in my CV.

Mantilla, about the omnipresence of mathematical ability and the unique ways every community uses it. The same is true for plant biology: every community and individual not only has a unique perspective of plants but has creatively innovated unique relationships and uses with them. To not consider all perspectives is a biased scientific result, and given the inequities of the academy, certainly our knowledge is at best incomplete or at worst distorted.

In this essay I will describe geometric approaches I am using to estimate and predict underlying latent spaces that guide the evolution and development of leaf shape. I will then discuss applying TDA to study the plant form, from the work of recently graduated PhD student Dr. Erik Amézquita who applied TDA-based approaches to X-ray CT scans of barley (Amézquita et al., 2021), citrus (Amézquita et al., 2022), and walnut, to a study analyzing the structure of gene expression profiles across the flowering plants from mathematically defined perspectives of development and stress (pre-print Palande et al., 2022). The latter work on gene expression was conducted as part of a bilingual and publicly available curriculum I developed called *Plants&Python* (VanBuren et al., 2022) that teaches coding using Python and is taught to students for credit at both Michigan State University (MSU) and Universidad Nacional Autónoma de México (UNAM). I elaborate on the class and pedagogical data more in my teaching portfolio but will highlight a bibliometric work recently published in the *Proceedings of the National Academies of Sciences* quantifying the effects of colonialism and patriarchy in the plant science literature (Marks et al., 2023). Just as topological theory tells us that the shape of data depends on how you look at it and that a single perspective is incomplete, similarly the exclusion of different cultural perspectives creates biased results, arising from and through the entire scientific process, from teaching and hypothesis generation to data collection and analysis. ***For me, bridging disciplines and cultures is a natural extension of the mathematics I use in my research and informs the intended impacts of the ways I integrate teaching and research and the communities I choose to associate with.***

Geometry and leaves: estimating morphospaces so that we can model and predict shape

I started my career by measuring the genetic (Chitwood et al., 2014; Demmings et al., 2019), developmental (Chitwood et al., 2016; Migicovsky et al., 2022), rootstock-derived (Migicovsky et al., 2019; Harris et al., 2021), and environmental (Chitwood et al., 2016; Baumgartner et al., 2020; Chitwood et al., 2021) effects on leaf shape, in diverse species, including grapevine (see references above), tomato (Chitwood et al., 2012a; 2012b; 2013; 2014; 2015; Coneva et al., 2017, highlighted in *National Geographic*[5]), maracuyá (passionvine; Chitwood and Otoni, 2017a; 2017b), cotton (Andres et al., 2017), *Alstroemeria* (Chitwood et al., 2012c), apple (Migicovsky et al., 2018), and more, including the cultural evolution of violin shape (Chitwood, 2014), highlighted in mainstream media, including *The New York Times*[6], NPR's *Science Friday*[7], *Los Angeles Times*[8], *El País*[9], and *Scientific American*[10]. The work on grapevine leaf shape was in part funded through an NSF Plant Genome Research Program (PGRP) grant to study the environmental regulation of rootstock-derived effects on the shoot (#1546869, Total award: $4,647,589, original sub-award: $926,896, portion of sub-award initially transferred to MSU: $657,169.47, amended to $603,845.18 on 4 Nov 2021). In all these cases I used geometric morphometrics, either Procrustean landmark- or Fourier-based methods, to measure effects using statistical techniques. But there is more: in measuring leaves across the flowering plants and locations

around the world (Li et al., 2018) larger questions, such as what leaf shapes are possible and how do they compare to the shapes that exist, arose. To answer these questions requires estimating the underlaying latent spaces that evolutionary novelty and developmental constraint are drawing upon.

My current work uses shapes (closed contours) as a simple case to measure these latent spaces, from which I derive continuous models of leaf shape as a function of development or evolution. Such models can be predictive and can capture the dynamic morphology of leaves that is missing when leaves are treated as static images. The latent spaces we are measuring in 2D are exactly the underlying structures of data that we aim to measure using TDA for higher dimensional data, and inform our efforts working on more complex examples, as described in the next section. As part of the 2019 *Plants&Python* class,

Figure 1: A high-resolution latent morphospace of grapevine leaves

student co-authors created continuous models of grapevine leaf development, which they showed better discriminated species than individual leaves (Bryson et al., 2020). Undergraduate Professorial Assistant Joey Mullins, as part of a Dept. CMSE class project, developed Python code to use the inverse Principal Component Analysis (PCA) transform to predict the underlying evolutionary and developmental morphospace of grapevine leaves (Chitwood & Mullins, 2022). Similarly, I gave students in the CMSE 201 class I teach the option to contribute to a research project as co-authors modeling tree rings of a 30-year-old rootstock trial vineyard we scanned using X-ray CT for their class project (Migicovksy et al., 2022). Building on a work paying homage to the late ampelographer (the study of ampelography identifies grapevines, with a geometric focus on leaf shape) Pierre Galet (Chitwood, 2020) I have been developing high resolution morphospaces for grapevine leaves (**Figure 1**) and extending these approaches using an intra-leaf modeling technique to normalize varying lobe number in *Cannabis* leaves with Visiting Graduate Student Manica Balant. The above modeling approaches share the aims of measuring data comprehensively and, rather than focusing on individual data points, estimate and visualize the underlying data structure, which are the exact motivations behind using TDA on more complex data.
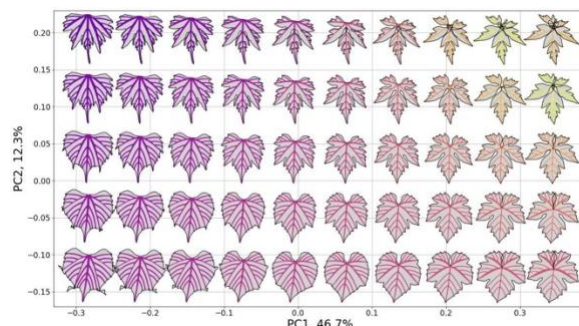
Topology and perspective: extracting information from shape and visualizing data structure
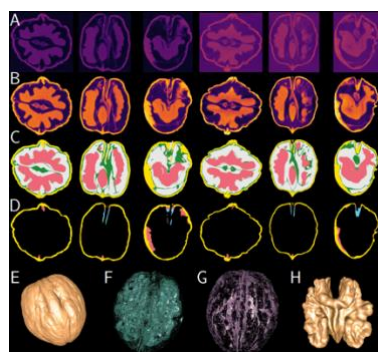

Figure 2: Image processing of walnut X-ray CT scans.

The above examples of leaf shape or other simple features (like tree rings) are easily modeled, but ultimately plants are more complex. Former graduate student Dr. Erik Amézquita, who graduated spring semester 2023 and will be starting as a Preparing Future Faculty—Faculty Diversity Fellow (PFFFD) at University of Missouri (Mizzou) in fall 2023, used image processing methods on X-ray CT scans to isolate specific tissues in walnut (**Figure 2**) from a breeding program in collaboration with Dr. Pat Brown (UC Davis) and similarly isolated tissues, focusing on the patterning of the oil glands, in citrus from the UC Riverside collection in collaboration with Dr. Danelle Seymour (Amézquita et al., 2022). To isolate plant tissues in 3D with such precision is a feat, but still the question remains: how to extract the information embedded in these intricate structures? For this question, geometric morphometrics is inadequate and topological theory is needed. The Euler Characteristic Transform (ECT) states that, given

an infinite number of axes through an object serving as lens functions, and an infinite number of threshold values along these axes, that through calculating the Euler Characteristic (a topological invariant feature) at each threshold of each axis, any object can be distinguished from any other[11]. Subsequent work has shown that a finite (but still undefined and large number of axes and thresholds) is sufficient for this task. In collaboration with Dr. Daniel Koenig (UC Riverside), Dr. Amézquita analyzed thousands of barley seeds from an artificial evolution experiment grown for nearly 60 generations resulting from crosses between the major barley varieties from across the world (Amézquita et al., 2021). Applying the ECT to founder lines, he then created a model on the parents and applied it to progeny in subsequent generations. Using the shape of seeds alone as measured using the ECT, the model predicted the presence of genetic material from parents that came to dominate these populations, and we are currently following up on this work in collaboration with Dr. Daniel Koenig.

The ECT aims to extract the total information content embedded in a data structure using large numbers of lens functions. But lens functions, as their name implies, can also selectively be used to look at data structures from defined perspectives. The same data has different shapes, depending on how you look at it. This was the goal of the 2020 *Plants&Python* class project (pre-print Palande et al., 2022), in which student co-authors visualized the structure of gene expression across the flowering plants as a graph, from the perspectives of development and environment. The actual data, a pairwise distance matrix of orthogroup expression profiles across flowering plants sampled from different tissues and stress conditions, remains constant. The lens functions, continuous models providing a real number value reflecting tissue identity or stress treatment, are then applied to each data point. These two lens function values are then used with a TDA algorithm known as Mapper[12] to visualize the data structure as a graph. The same data

Figure 3: Graphs of gene expression in flowering plants as viewed through tissue (top) and stress (bottom) lenses.

(gene expression) has two distinct structures depending on how you view the data, either through the lens of tissue identity or stress response (**Figure 3**), revealing conserved and diverging molecular mechanisms that influence these two distinct identities within plants.
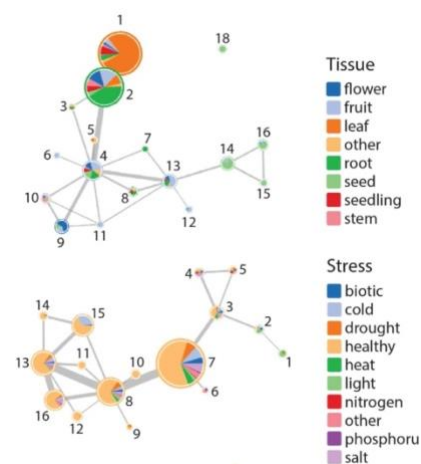
Through a new award from the NSF PGRP (#IOS-2310355, Total award: $1,500,000; $639,998 to MSU; 1 June 2023 – 31 May 2026) **"Collaborative Research: RESEARCH-PGR: Predicting phenotype from molecular profiles with deep learning: using Topological Data Analysis to address a Grand Challenge in the plant sciences"** we build on the above use of TDA to estimate the underlying data structures of both phenotypic and molecular spaces. In collaboration with Dr. Aman Husbands (University of Pennsylvania) and Dr. Robert VanBuren (Dept. Horticulture) we will isolate *in silico* individual leaves in phyllotactic order from X-ray CT scans of *Arabidopsis* ecotype rosettes that vary in developmental robustness. From the same individual leaves, we will create RNA-Seq expression profiles. In collaboration with Dr. Elizabeth Munch (Dept. CMSE) we will create Mapper graph representations of the underlying structures of corresponding phenotypic and molecular profile spaces, using a developmental lens based on node number. Using these graph-based representations, in collaboration with Dr. Arjun Krishnan (formerly

---

[11] Turner, Katharine, Sayan Mukherjee, and Doug M. Boyer. "Persistent homology transform for modeling shapes and surfaces." *Information and Inference: A Journal of the IMA* 3.4 (2014): 310-344.

[12] Singh, Gurjeet, Facundo Mémoli, and Gunnar E. Carlsson. "Topological methods for the analysis of high dimensional data sets and 3d object recognition." *PBG@ Eurographics* 2 (2007): 091-100.

Dept. CMSE, now University of Colorado, Anschutz) we will apply Convolutional Neural Networks (CNNs) on graph embeddings to create a model predicting phenotype from molecular profiles, a grand challenge in the plant sciences.

The different perspectives we bring to science

The novelty of my work is mostly metaphorical: that the way we measure and interact with plants—either through their morphology or through molecular profiles—is a type of data, and therefore techniques from computational mathematics, including TDA and the knowledge that perspective influences data structure, can be applied. The concept of perspective influencing perceived reality is not new, and the concept of "lens" in the humanities (as in epistemic perspective) predates the use of filter ("lens") functions in TDA. Just as the structure of data depends on the mathematically defined perspective from which we view it, so too does reality itself depending on our background, and therefore the axioms of Frederico Ardila Mantilla that open this essay are not only a null hypothesis against which we test our professed values in science that we too often assume to be self-evident, but a warning that science—knowledge—is socially constructed and that to deviate from these axioms not only is unethical, but results in an incomplete and distorted world view.
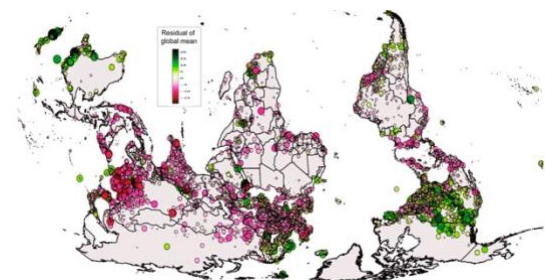

Figure 4: Global plant science citations adjusted by population.

Together with *Plants&Python* instructors Dr. Robert VanBuren in Dept. Horticulture and Dr. Alejandra Rougon Cardoso (UNAM ENES León) we critically analyzed over 300,000 citations from the last 20 years in plant science (Marks et al., 2023). The work demonstrated profound geographical bias that correlated with national affluence (**Figure 4**) and gender biases inferred from masculine- and feminine-associated names. Through bibliometric analysis, the work quantifies oppressive forces of colonialism and patriarchy that we know operate to create global inequities in the plant sciences. Although not a complete solution, the sharing of intellectual capital and technical skills sequestered by the Global North would be one way to equalize these disparities. ***Please see my teaching portfolio for more details***, but together with Drs. VanBuren and Rougon Cardoso, I developed the *Plants&Python* curriculum, a freely available[13] and bilingual set of lessons that teach coding in Python from scratch using plant-inspired examples. The class is taught in hybrid format and for credit at MSU as HRT841: *Foundation in Plant and Computational Sciences* and throughout México in the UNAM system as *Temas selectos: bioinformática y minería de datos con python*. Not only is it a vehicle to teach coding to biologists in an international setting, the class is also a way to conduct cutting-edge research, and each year we publish a manuscript with student co-authors (see Bryson et al., 2020 and pre-print Palande et al., 2022). Large biological datasets are ideally suited for a classroom setting and a collaborative, critical environment where large numbers of perspectives are considered. As we know from TDA, to not consider the structure of the shared reality we are measuring from multiple viewpoints is lacking and incomplete. It is possible to leverage diversity—which follows from equity and justice—to practice a sounder science. ***Going forward, I intend to continue building on the multilingual and international teaching platforms I have developed thus far to include students with diverse talents and contributions in large-scale analyses of the underlying data structures that shape the biological world around us.***

---

[13] https://plantsandpython.github.io/PlantsAndPython