Understanding Processing-Structure-Mechanical Property Relationships in Sustainable Biopolymer-based Composites Using Design of Experiments and Machine Learning

Jeremy L. Fredricks

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2023

Reading Committee:
Eleftheria Roumeli, Chair
Miqin Zhang
Navid Zobeiry
Renata Bura

Program Authorized to Offer Degree:
Materials Science and Engineering

University of Washington

Abstract

Understanding Processing-Structure-Mechanical Property Relationships in Sustainable Biopolymer-based Composites Using Design of Experiments and Machine Learning

Jeremy L. Fredricks

Chair of the Supervisory Committee:
Professor Eleftheria Roumeli
Materials Science and Engineering

Nature uses hierarchy as a means to imbue materials with mechanical and physical properties that are greater than the sum of its parts. Molecules are combined into polymers which bundle together into fibers that make up ever-larger structural building blocks that comprise the organism. Wood exemplifies this concept, with cellulose as the main polymer backbone supported by biopolymers such as hemicellulose, pectin, and lignin. These components act as binders to toughen the tree against environmental forces. Moreover, the resulting composite exhibits functional features, such as the ability to transport nutrients throughout the structure.

One of the most common approaches to making synthetic biocomposites has been through the extraction and use of select components from biomass. However, despite the use of the same materials used by nature, the properties observed in natural composites have proven difficult to replicate. As an alternative approach to making better biocomposites, there has been a research push in recent years toward exploring the use of the entire organism as a component. The recently

coined term “biomatter” refers to such materials wherein the entire biological matter is used without removing any components, retaining native or minimally altered hierarchical nano- and microstructures. Biomatter materials offer a promising solution to address the need for sustainable composites; none of the components within the material go to waste or lead to waste from inefficient extraction processes, and the mechanical properties have already been optimized against stressors by the hierarchical design inherent to natural materials. In addition, by shifting away from petroleum feedstocks and harnessing abundant, renewable biological resources, we can fabricate materials that align with sustainability goals.

Another approach to making better biocomposites has been to combine natural materials not normally found together in nature. By leveraging general chemical and physical principles, best-in-class materials can be combined to rationally design composite materials. However, we currently lack the knowledge to construct composites that maximize the utility of materials with intricate and dissimilar components. The work presented in this thesis tackles this issue by employing three different strategies to rationally build composites with biomatter building blocks.

We started by selecting model materials of varying complexity, ranging from molecules to polymers to entire cells, as composite components. For the cell building blocks, we opted for microalgae, specifically *Spirulina platensis* (“spirulina”) and *Chlorella sp.* (“chlorella”). These algae offer several advantages, including their abundance in industries such as biofuel, pharmaceuticals, and nutraceuticals, ease of culturing and growth, absence of complex tissue-like structures, and a diverse polymer composition of proteins, short and long polysaccharides, phenolics, and small molecules, all of which serve as valuable components for constructing polymer-based composites. We selected bacterial cellulose (BC) fibers grown from bacteria-yeast cocultures as a model structural polymer. BC fibers possess a high molecular weight, a high degree of crystallinity, and exceptional mechanical properties, owing to their inherent hierarchical structure. Notably, BC fibers are the sole product of this culture system, eliminating the need for separation and extraction after harvesting. Lignin, an abundant phenolic polymer found in woody biomass and a byproduct of the pulp and paper-making industries, was chosen to serve as a binder in our composite systems. Lignin's rich aromatic backbone and diverse functional groups allow for secondary interactions with glucans, like cellulose, making it an appropriate choice as a binder. Lastly, we incorporated stearic acid as a small, functional molecule. Stearic acid is a native component of many types of biomass, and its amphiphilic structure enables it to function as a plasticizer and/or surfactant within the composite system.

The three strategies chosen to guide the development of multiscale biomatter composites were a traditional trial-and-error approach, design of experiments, and machine learning. Using the trial-and-error approach, we investigated the effects of three different post-processing methods and the influence of plant-extracted micro-crystalline cellulose fibers (CFs) on the structure and mechanical properties of a spirulina-based matrix. We aimed to establish methods by which a single biomatter component could be used to create multiscale objects. The smallest scale consisted of the polymeric composite units inherent in the cell walls of spirulina. The next scale consisted of individual cells of spirulina. The subsequent scale consisted of physically-confined particles of spirulina chains. The final scale was dictated by the macroscale-controllable geometry extruded via a 3D printer.

We employed a design of experiments approach to explore how processing conditions influence the *ex-vivo* incorporation of lignin into a BC matrix in hierarchical lignocellulose papers. We used heat, pressure, and pressing time as parameters in a hot-pressing process to facilitate a systematic understanding of the processing-property relationships. Our analysis revealed optimal processing temperature/pressure/time conditions for tuning mechanical and water-repellency properties via structural changes in the processed papers.

Finally, we developed machine learning models based on a Bayesian Optimization approach to guide the search for optimal compositions of ternary composites for improved mechanical properties. Chlorella, BC, stearic acid, and a water-ethanol solution were combined, processed, and cast into films to facilitate understanding of the relationship between composition and the resulting mechanical properties of the composites.

We conclude with a summary of the benefits and disadvantages of using traditional experimental approaches, statistical designs of experiments, and machine learning models to design and fabricate next-generation, sustainable biocomposites.

Acknowledgements

I thank the Royalty Research Fund, Microsoft Research, and Meta Labs for their generous pecuniary support in my research. I thank the committee members of my qualifying, general, and final examinations for their generous intellectual support and time: Drs. Navid Zobeiry, Dwayne Arola, Renata Bura, Lilo Pozzo, and Miqin Zhang. I thank my PI, Eleftheria Roumeli, for her generous pecuniary, intellectual, and emotional support, and her everlasting patience.

Working with Eleftheria has been a unique and transformative experience in my life. Not quite a peer friend, but more caring or understanding than a boss; not as much hands-on guidance as a mentor, but more amicable than an advisor; the best way that I can describe my role under Eleftheria is as the shadow under a leader. She was always encouraging with her words, but I was actually inspired, motivated, and moved to action because I could see that they were backed by her own personal experiences and strong work ethic. Through the years I've known her, she always did her best to adapt to and accommodate me and others. I am truly grateful to have had for a PI someone I could genuinely look up to and seek for personal inspiration when the need arose.

In my mind, one of Eleftheria's unique strengths is her ability to pick good people. All of the postdocs and PhD students who joined after me in the lab have contributed to the lab culture in their own, idiosyncratic ways, synchronizing as a collective despite such differences in personality and work styles. I want to acknowledge that their presence in the group enriched the latter half of my PhD experience in a way I did not know that I needed in my life. I would like to thank the postdocs Drs. Andrew Jimenez and Paul Grandgeorge. Interacting with them was the closest I got to having mentors in the lab. I was able to glean from them a great deal of knowledge that only postdocs would know; from unconventional and utilitarian lab tips and tricks to work-life balance as a scientist and person, they filled a niche in my learning experience that I would otherwise have not known I was missing. I would like to thank my fellow PhD labmates, in order of distance from my desk: Hareesh Iyer, Julia Didier, Meng-Yen Lin, Tim Liao, Mallory Parker, and Ian Campbell. Their kindness, intelligence, scientific prowess, and passion for Dessert Mondays helped keep me grounded during the ups and downs of my time in lab. There are too many good things that I could and want to say about them and their positive impact on my life, but the length of this dissertation would double.

I want to thank my family for background support. They usually didn't understand what I did, but the love and support was there. In particular, I would like to thank my grandfather, Hiroaki Sato, for sparking my interest as a scholar since I was young and paving the way to my becoming a scientist. I would like to also thank my fiancée, Catherine Tong, for being my backboard for good and bad ideas, and listening to my complaints and bouts of jubilance all the same. She helped me keep the rest of my life in balance when labwork became overwhelming.

Finally, I would like to thank the people outside of my research and family that helped establish the "balance" part of my work-life balance: the coaches and students I was able to meet through taekwondo because I was a student, and the teachers and classmates I've had in my business classes, competitions, and endeavors. They helped me keep in perspective that lab was only a part of my life, but not life itself.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

Despite their ubiquity and usefulness, plastic materials have fallen under scrutiny in recent years due to their detrimental environmental and health effects. Their environmental impacts are exacerbated by the fact that major developed regions, such as the US, Europe, Australia, and Japan, classify plastics as generic solid waste – a classification that usually places them in landfills and eventually oceans at end-of-life with no opportunities to break down into more benign substances.[1] To circumvent this issue, biodegradable plastics, whose properties have allowed supplantation of some general-purpose plastics, have been developed. Examples of these plastics include polylactic acid (PLA), polycaprolactone (PCL), polyhydroxyalkanoates (PHAs), polybutylene succinate (PBS), and polybutylene adipate terephthalate (PBAT).[2] Despite being biodegradable, these plastics were designed to degrade in aerobic environments, which general landfills are not.[3,4] As a result, biodegradable plastics still tend to degrade more slowly than they are produced and consumed.

1.1 Strategies toward end-of-life plastics management

Because there are not yet enough options for completely replacing useful but nondegradable plastics, steps have been taken to mitigate their post-use impact. For example, strategies have been developed for (1) making difficult-to-recycle aromatic polymers easier to break down, (2) reusing recycled plastics via addition of compatibilizers, and (3) adding dynamic cross-links to traditionally nonrecyclable thermosets to more easily degrade them.[5] These strategies are explained in more detail next and summarized in **Fig. 1.1a**.

Some commonly used commodity polyesters such as polyethylene terephthalate (PET) are aromatic, meaning they contain cyclic rings in the backbone that provide rigidity and strength to the plastic. Unfortunately, the ester bonds comprising such backbones are usually difficult to break apart. Efforts are being undertaken to make aromatic polymers easier to degrade by introducing easier targets into the backbone; by adding aliphatic (non-ring) polyesters, such as PLA, to the backbone of the polymer, more easily hydrolysable ester bonds are introduced.[5] **Fig. 1.1e** illustrates such an addition that facilitates the decomposition of an aromatic polymer. Recently, polymerization techniques have also been developed that allow the conversion of aromatic ester molecules (lactones), commonly found in plant matter, into aliphatic polymer chains through ring-opening reactions (**Fig. 1.1b**).[5,6]

Mechanical recycling is the primary method of conventional, industrial-scale plastics recycling.[7] This process starts with the sorting of plastics by type and color because most plastics are immiscible, phase separating if they are directly melted and reprocessed together, resulting in brittle or nonuniform plastic blends.[5,7] This immiscibility poses problems for the recycling process when plastic feedstocks are contaminated with even small amounts of impurities. One way to circumvent this issue is through the addition of a compatibilizer that lowers the interfacial energies between the immiscible components by containing fractions of thermodynamically compatible components of each phase in the melt in the form of a block or graft copolymer.[5,7] Literature has shown that this general strategy works for a variety of polymer blends.[5] Compatibilizing increases the probability of reuse of plastics at their end of life as the compatibilized polymer blends show properties that are often in between that of the parent polymers. **Fig. 1.1c** illustrates an example of a phase-separated blend and a compatibilizer-added blend. **Fig. 1.1d** represents the potential mechanisms by which block polymers may compatibilize an immiscible blend.

Thermosets are a class of polymer that cannot be melt processed due to the presence of cross-links in the polymer structure.[5] Polymers with dynamic cross-links that preserve the properties of traditional thermosets but allow degradation, or re-cross-linking and thus reprocessing/recycling, are being developed. Dynamically cross-linked polymers work under the same use conditions as traditional thermosets, but become uncross-linked when subjected to specific reprocessing conditions (e.g. an acidic environment, introduction of a specific degradation catalyst, heat).[5,7] **Fig. 1.2** illustrates the added functionality introduced by dynamic cross-links.

Enzymatic approaches for depolymerizing plastics have been investigated as well. For PET, which represents 9% of all plastics produced,[7] there are at least 20 PET-hydrolyzing enzymes that have been developed; unfortunately, few show appreciable activity under ambient conditions.[8] A promising study showed that machine learning methods could be used to design enzymes that have higher degradation activity than wild-type enzymes, at temperatures as low as 30 °C and in a functionally large pH range (6.5 – 8.0).[8] **Fig. 1.1f** illustrates the circular lifecycle of a piece of PET plastic degraded into its monomer by the designed enzyme and then repolymerized again as a film.

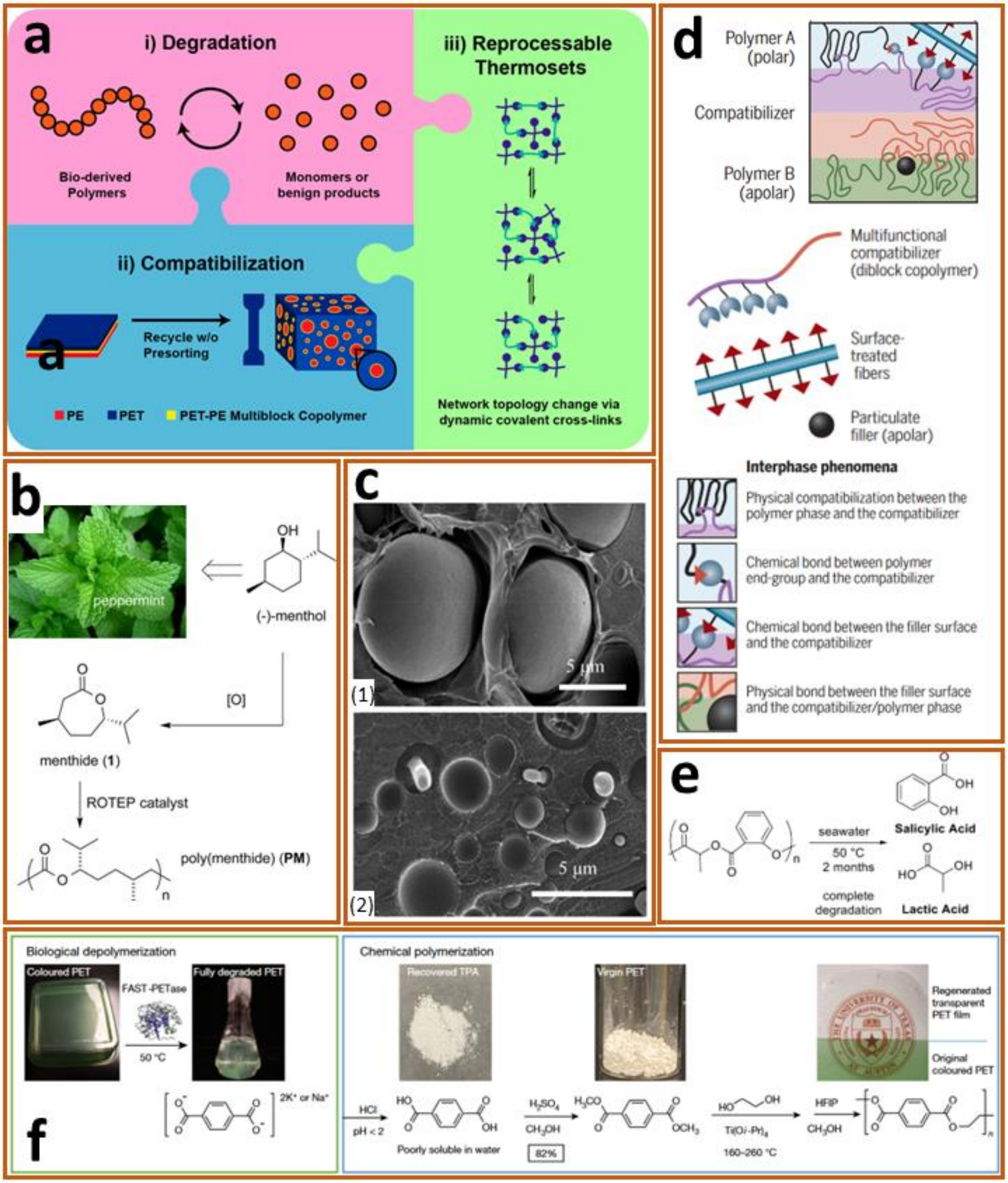


Figure 1.1:

(a) Strategies for mitigating post-use impact: i) Designing polymers that are easy to degrade into their monomeric form after use; ii) Introduction of compatibilizers for mixed plastic waste; iii) Development of thermosets with dynamic cross-links that are essential for giving new shape to the plastics after use. (b) The addition of an aliphatic chain into an aromatic polymer may facilitate its decomposition into benign products via introduction of a facilely degradable ester bond. (c) A

lactone (menthol) derived from natural peppermint and then the use of a ring-opening polymerization technique that attacks the ester bond in the molecule to convert it into an aliphatic polymer. (d) SEM images of cryofractured PE/PET blends (1) without and (2) with a compatibilizer present. Interfacial adhesion is better with compatibilizer while there are visible air gaps present after fracturing in the uncompatibilized blend. (e) Possible functionalities added to compatibilizing block copolymers to improve compatibility with immiscible parent polymer blends. (f) Degradation of a colored piece of PET plastic by a designed enzyme. The monomers can be repolymerized into a new piece of plastic.
(a), (c), and (e) reproduced/adapted from Haque et al.;[5] (b) reproduced from Hillmyer et al.;[6] (d) adapted from Mohanty et al.;[2] (f) reproduced from Lu et al.[8]

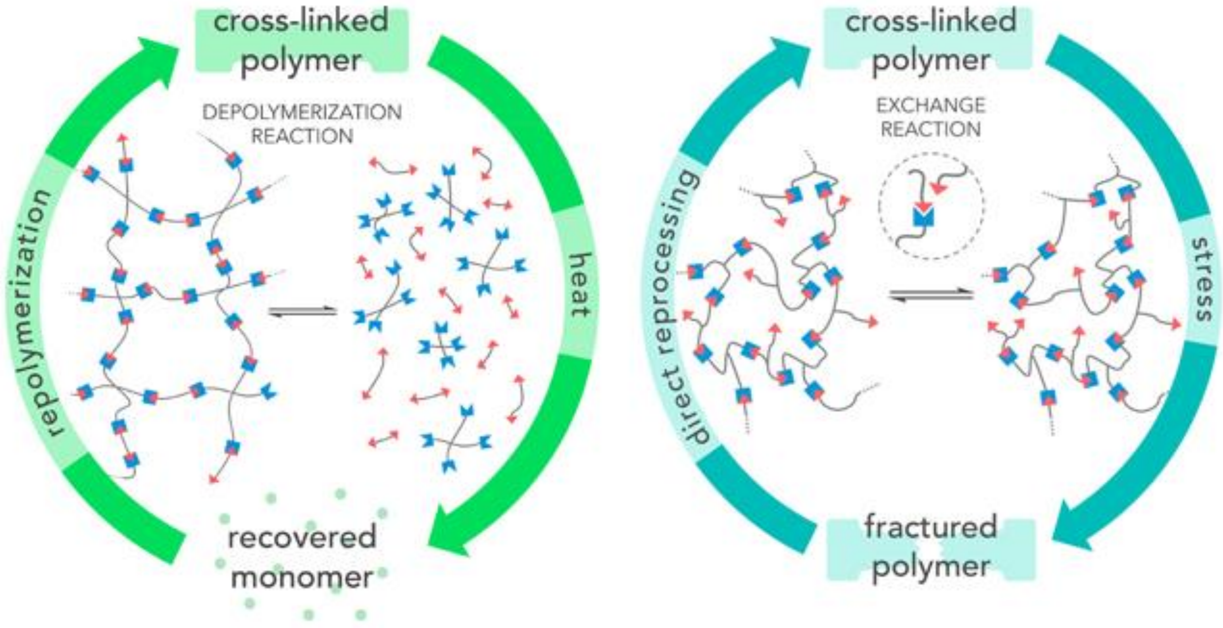


Figure 1.2: Physical and heat-induced recovery facilitated by the addition of dynamic cross-links to thermosets. Dynamic cross-links may give rise to properties such as degradability into the monomeric state or self-healing (i.e. healing without external additives).
Reproduced from Fortman et al.[9]

1.2 Monomer extraction from biomass

Designing polymers with end-of-life in mind to prolong their use is an important step toward a more sustainable society. However, sourcing of polymers from materials that already exist, such as waste products from other industries, is another approach to creating a circular economy. One of the approaches toward reducing environmental impact of plastics is to extract or derive plastic monomers from biological matter (biomatter) rather than petroleum sources. While the impacts of the final use are not mitigated, there are potentially life cycle benefits for changing sourcing. For example, a study found that low-density polyethylene (LDPE) sourced from sugarcane consumed more energy initially to produce but had smaller global warming impact compared to oil-based LDPE when taking into account both of their sourcing, the land required to source, refining, polymerization, and shipping.[2,10]

Sourcing from biomass can be more expensive than sourcing from oil deposits, especially for commodity monomers/polymers, due to processing infrastructure that is already in place for oil-based feedstocks. Biomass may take many more steps than petroleum-based feedstocks to convert from their raw form to their intended monomer form, leading to inefficiencies in production.[5] To solve this issue, developing highly active and selective catalysts for conversion of biomass-derived chemicals has been a fruitful area of research.

Another approach that is being researched is the development of new monomers from biomass with comparable properties to more common monomers. While full conversion from biomass to a specific monomer may take many steps, a chemical intermediate may also yield similar or useful properties that would be otherwise difficult to attain synthetically. For example, the inclusion of heteroatoms, specific stereochemistry or molecular topology, or reactive moieties may be benefits inherent in the biomass starting molecule(s).[5,11]

1.3 Extraction of ready-to-use polymers directly from biomass

Deriving monomers from biomass is one approach to obtaining building blocks for sustainable materials. Another approach is to source entire polymers from biological matter. These polymers are generally obtained from living organisms that produce the polymers as-is, such as in the case of cellulose from some species of bacteria, or in a form that can be easily attained through minimal processing of a natural material, such as in the case of lignocellulosic materials from wood and wood products. The benefits of extracting whole polymers instead of monomers include fewer required processing steps, hierarchical morphologies that would be difficult to attain otherwise, and (an)isotropic structures/properties that are naturally present as a result of their natural origins.

Using cellulose as a material is a rapidly growing field due to cellulose's potential as a sustainable plastic alternative with its inherent strength, abundance, and ability to be readily extracted from renewable biomass and processed into new products.[12–18] The oxygenated molecular chains comprising cellulose allow for its degradation by bacteria, fungi, and yeasts present in soil while its hydroxyl groups allow for chemical functionalization by established chemical means.[13] Most commonly, cellulose is extracted from woody biomass in a process that involves chemically stripping lignin and other cell wall components from the biomass and then processing the remaining material.[12,17] Processing, in this case, refers to both treatments to purify the cellulose and fibrillation that may occur in the course of removing the other components. Because wood is a naturally hierarchical material, wood cellulose is present as fibers and fibrils at different length scales. Fibrillating cellulose involves reducing the sizes of the fibrils to less than 100 µm to as low as 3 nm.[13] Size reduction in cellulose is desired due to its anomalous strengthening and toughening behavior as fibril size decreases.[19] Furthermore, fibrillating cellulose presents opportunities to confer anisotropy by aligning fibrils and introducing or modifying porosity in films, coatings, and membranes of cellulosic materials.[13]

Recently, studies have begun to focus on lignocellulosic materials, which take advantage of the upwards of 30% lignin present in woody biomass for potential improvements in strength and hydrophobicity in the final product.[20] One option to modify properties of lignocellulosic materials is by changing the cellulose source.[12,21] For example, a highly crystalline, pure cellulose nanofibrillar matrix can be naturally synthesized by certain types of bacteria in the form of a layered, interconnected stack of individual cellulose sheets (referred to as “pellicle”). Bacterial cellulose (BC) pellicles circumvent the unavoidable cellulose matrix degradation that occurs during wood delignification treatments.[22]

Two different strategies for creating lignocellulosic materials are by introduction of lignin as a filler in pure cellulose matrix materials, and the removal of other cell wall components from wood to leave a lignocellulose framework. The former strategy allows for any type of cellulose to be used as a matrix and, as a result, has been applied to introduce lignin in wood-extracted micro- and nanocellulose blends.[15,23,24] Wang et al. demonstrated that heat-treating of cellulose-lignin blends allows lignin to soften and fill the voids of a cellulose network, improving mechanical properties and water stability in the resultant composite.[15] Prior results from Jiang et al. concur with these observations and provide evidence that altering hot-pressing temperature modulates the mechanical properties of the lignocellulosic composite.[24] Specifically, the reintroduction of lignin back into totally delignified cellulose matrices, at optimal press temperatures (100 - 130 °C), showed a 5 times higher tensile strength in the composite compared to a conventional cellulose paper.[24] In another study, Farooq et al. introduced different amounts and types of lignin into nanofibrillated cellulose (from wood pulp) and reported a drastic improvement on the toughness of cellulose composite using 10 wt.% colloidal lignin particles. They also reported an increase in the stiffness of the composites at the expense of toughness when they hot-pressed their papers at 100 °C.[25]

More recently, Xia et al. created a lignocellulosic composite from wood chips by dissolving the lignin contained in the biomass and then reconstituting it. The natural structure of the cellulose micro-/nanofibrils matrix, which was originally held together by the lignin binder, consequently fell apart. After lignin reprecipitation, a slurry of dispersed cellulose in a lignin matrix was attained.[16] Interesting, the solvent that they used to dissolve the lignin also fibrillated the cellulose into widths of 10-300 nm. The regenerated lignin was also altered to be amphiphilic, its hydrophobic component conferring water stability to the film while the hydrophilic moieties increased interactions with the cellulose and improved the strength and toughness 7- and 8-fold, respectively, over a cellulose-only control film from the same wood source. This study built on a prior study by Song et al. who reported that partial wood delignification (via chemical treatment followed by hot-pressing) can lead to self-bonded composites with closed pore structures with enhanced strengths, which were maximized at 10 - 15 wt.% lignin.[26] Finally, Xiao et al. dried and rehydrated partially delignified wood biomass to make a densified structure that was stronger than its natural wood counterpart. The drying and rehydrating process effected a different microstructure compared to only partially air-drying to the same water content. This difference in microstructure gave rise to a difference in failure mechanisms, whereby the reswelling process prestressed the wood microstructure so that the cell walls could sustain repeated compressive and tensile stresses from folding whereas partially air-dried samples brittlely failed after folding, demonstrating the importance of processing on structure and performance.[27]

1.4 Utilization of the entire biomass and processing of whole-cell, whole-organism biomatter

A third approach to creating sustainable materials is to utilize the entire biomass without purification or extraction of components. There is overlap with traditional biocomposites in this approach, introducing biomass as filler in synthetic plastics. Specifically, combinations of chlorella microalgae with polyethylene,[28] spirulina microalgae with poly(butylene succinate),[29] ulva macroalgae with polyvinyl alcohol[30] or polypropylene[31], and various macroalgae with polylactic acid (PLA)[32] have been reported. However, materials made primarily or entirely of organisms/tissue have also shown promise, particularly with algae-based materials.[33–35] For example, bacteria-algae symbiotic biocomposites have been created where algal cells direct the growth of BC in a biofilm matrix consisting of cells and cellulose.[36] Bioplastics of pure chlorella and spirulina biomass[37] and chlamydomonas biomass plasticized with glycerol[38] have been reported as well. In this report, bionanocomposites of spirulina were created and is elaborated in a later section.[39] Finally, biomatter nanocomposite materials made solely of cultured plant cells have been reported as well.[40]

In summary, fully-natural composites, and biomass-based materials (*biomatter*) in particular, have the potential to solve sourcing and degradation issues with their main benefits being that they are renewable, compostable, and contain biopolymers which can be used without further processing.[41,42] Furthermore, structural or hierarchical features that are difficult to mimic synthetically are present in natural sources,[43] enabling tunability of performance by source selection, processing strategies that can controllably alter structural features (e.g. by inducing fiber alignment), or introduction of fillers.

1.5 Specific processing methods for biomass composites: 3D-printing and lyophilization

The use of 3D-printing, or additive manufacturing, has gained traction in recent years for the relatively low starting cost of obtaining a printer, the promise of repeatable, customizable, and high-resolution prints, and compatibility with a variety of materials. The most commonly used materials in extrusion-based 3D-printing are synthetic plastics, such as acrylonitrile butadiene styrene (ABS) and PLA in the form of filaments.[44–46] Natural fibers, such as hemp, which are composites of multiple biopolymers, and extracted single-component polymer fibers, such as cellulose, are often used as fillers in synthetic plastics.[3,47] Such biocomposite materials have been widely studied as filaments for 3D-printing in the past decade.[48,49] The poor adhesion between natural fibers and synthetic plastics, difficulties in overcoming fiber agglomeration to achieve optimal dispersion, along with the innate hydrophilicity of natural fibers limit the applicability of such biocomposite materials to 3D-printing.[48,49]

A recent study showed that entirely biobased, amylopectin/cotton fiber mixtures can be used as standalone 3D-printing materials, and the produced structures can achieve properties comparable to commonly used plastic filaments upon post-processing.[50] This finding opens up new possibilities for integrating fully biomass-based materials (BBM) in extrusion-based 3D-printing, sometimes referred to as Direct Ink Writing in this context. BBM can be extruded while their components are at ambient temperature, which is a condition not met by plastic feedstock that require initial melting of stock filament. To be able to flow under ambient conditions, BBM are mixed with a solvent, most often water. Thus, a final drying step is required after printing to remove the added solvent and allow the printed structures to become structurally stable. Drying can be achieved through processes ranging from low-temperature techniques like lyophilization to more aggressive techniques such as microwave processing.[51] The drying post-processing step affects the final structure of the printed sample, influencing its density, shrinkage, and micro-morphology.[52–54] Thus, opportunities to utilize the dehydration process in order to control the structure and, consequently, mechanical properties of 3D-printed materials, emerge. The ability to tune the achieved mechanical properties would effectively expand the range of applications that these materials can reach.[55,56]

Commonly applied post-processing strategies for a polymer fabricated through extrusion-based printing include: (a) drying at various flow rates, pressures, and temperatures,[53–55] (b) heat treatments, either to promote thermal annealing[57] or gelatinization[50,58], and (c) exposure to select ion solutions to induce gelation or cross-linking[55,59]. Among drying processes, freeze-drying (lyophilization), which dehydrates materials through ice sublimation, is notable for allowing shape retention even in complex structures.

Depending on the freezing pattern, solvent, and sample composition, structures with different porosities and pore morphologies can be attained.[60–62] The main benefit of lyophilization is that it minimizes shrinkage and drying stresses that lead to warping and crack formation.[62] However, freeze-drying is slow and energy-intensive due to the low-temperature and -pressure requirements, and leads to inherently porous samples (aerogels).[55] In contrast, drying can be achieved quickly at ambient pressures and higher temperatures through oven-drying. Besides driving water evaporation, drying biomass-based materials at temperatures close to their gelatinization transition can effectively alter their micro-morphology by using the bound water to plasticize the biomass matrix.[50,58,63] The drawbacks of drying at elevated temperatures include the potential to damage and collapse the printed structure due to the morphology-dependent stress gradients produced within a material as its surface dries faster than its core.[64,65] Finally, drying methods which use ambient temperatures also exist. A closed desiccator containing a desiccant uses the humidity difference between the atmosphere and saturated material to drive evaporation. This method is a variation of a drying method known as low temperature, low humidity (LTLH) drying,[66] which has been used to dry hydrated structures while densifying their microstructure.

1.6 Biomass material composition selection using machine learning

Machine learning (ML) has come to the fore in the materials science field due to its promise of being able to draw insights from vast amounts of complex materials information that may be impossible for human researchers to connect manually.[67] ML methods have been successfully applied to areas of research where high-throughput methods are possible, to rapidly improve materials properties through data-driven insights.[68,69] However, a primary difficulty with using ML in materials science is lack of data due to often time-intensive or costly experiments resulting from the inability of common manufacturing methods to produce large datasets ($n > 50$) within a reasonable timeframe.[70–73] The difficulty of producing an accurate ML model using small datasets ($n < 50$) has resulted in the slow uptake of ML in fields where collecting experimental data can take weeks or months to complete. Furthermore, these same fields often encounter confounded or otherwise uncontrollable noise variables in their experimental designs. Combined with multi-step fabrication processes that introduce additional, albeit controlled, system variables, selecting informative features (i.e. defining unique input/output parameters for the model) becomes difficult.

High-throughput experimentation (HTE) was proposed as a means to conduct many small-scale experiments that use routine procedures. HTE uses a combinatoric approach hybridizing automation with arrayed or continuous designs and has been used in chemical process development,[74] particularly in pharmaceutics,[75] materials science,[76,77] and bioengineering[74]. However, this methodology may produce an overwhelming number of unnecessary results, being that every combination is tested. Several methods have been proposed to reduce the number of unnecessary experiments while achieving reliable insights from the limited data. Design of experiments (DOE) is a robust field for which many statistical methods and designs have been documented to gain insights or to drive the next steps in a study.[78–80] Additionally, ML methods have previously been combined with other computational methods, such as density functional theory (DFT) and finite element analysis (FEA), to help produce data to be used in a model; the use of these methods additionally produces more explainable models or reveal insights into how certain model parameters impact the model.[69,70] Expert-informed domain knowledge of physical, chemical, and structural parameters or descriptors have also been incorporated into ML algorithms to improve accuracy of a model by giving “starting points” for a model to base its analysis off of.[81,82] For example, the addition of relevant data points derived from less precise methods or procedures[71], or a transformation and use of descriptive categorical data with continuous data,[83] have been shown to improve accuracy of a model more than the addition of extra features (degrees of freedom) in an ML model might be able to achieve alone.

However, usually, researchers adjust the base algorithms or the machine learning workflow to adapt ML to their materials. Non-exhaustively, ML has been used to find optimal compositions in metal alloys[70,84] and glass[69] to optimize certain properties, in concrete to minimize greenhouse gas emissions while maximizing compressive strength,[85,86] with battery materials to predict battery lifetime based on initial charge-discharge behavior and then suggesting charging protocols that maximize the lifetime,[68] to predict and iteratively suggest optimal designs for composite materials,[87–89] to predict the stress-strain behavior of said composites based on the composite design,[90] in polymer nanocomposites to predict certain mechanical properties of polymer-nanofiller systems based on bulk property and microstructure information,[91] in other polymer systems to predict stress-strain behavior based on synthesis information (degree of conversion),[92] and in designing metamaterials.[93,94] Finally, ML methods have also been adapted for multi-objective optimization[80,85,86,95], that is, the balancing of tradeoffs and finding optima for multiple properties simultaneously given several constraints.

1.7 Thesis Overview

The aim of this dissertation is to compare and demonstrate different approaches to the rational design of materials. A traditional trial and error approach, design of experiments, and machine learning are utilized to guide the development of multiscale, biological matter (*biomatter*) composites and rationalize their mechanical property performances in the context of the interactions between their dissimilar components and processing parameters. The following is an overview of this dissertation's chapters.

In Chapter 3, we investigated the effects of three different post-processing methods and the influence of plant-extracted micro-crystalline cellulose fibers (CFs) on the structure and mechanical properties of a spirulina-based matrix. The ultimate goal of this project was to expand the attainable mechanical properties of the proposed fully biomass-based materials. We aimed to establish methods by which a single bionanocomposite component could be used to create multiscale objects. The smallest scale consisted of the polymeric composite units inherent in the cell walls of *Spirulina platensis* (hereafter denoted "spirulina"). The next scale consisted of individual cells of spirulina. The subsequent scale consisted of physically-confined particles of spirulina chains. The final scale was dictated by the macroscale-controllable geometry extruded via a 3D printer.

In Chapter 4, we explored how the incorporation of lignin, used as a natural binder, in bacterial cellulose (BC) papers affected the mechanical properties of the resulting 2-component, composite lignocellulose papers. The papers were processed under various conditions to enable tuning of the mechanical properties. We used heat, pressure, and pressing time as parameters and employed a design of experiments (DOE) approach to facilitate a systematic understanding of the processing-structure-property relationships when lignin particles are incorporated into the multiscale (nano- to centimeter) BC matrix. Our analysis revealed optimal processing temperature/pressure/time conditions and the overall effects of lignin infusion into the BC papers.

In Chapter 5, we developed machine learning models to optimize the composition of 3-component composites for improved mechanical properties. Chlorella, BC, stearic acid, and a water-ethanol solution were combined and processed into composite films to facilitate understanding of the relationship between the composition and the resulting mechanical properties of the composites.

2. Materials and Methods

All materials and methods used throughout the subsequent chapters are compiled here.

2.1 Overview of Materials used and Methods

2.1.1 Characterization Techniques

SEM, TEM, Optical Microscopy, XRD, FTIR, TGA, DSC, HPLC, Rheological Analysis, Water Contact Angle, Stylus Profilometry, Viscometry

2.1.2 Mechanical Testing Techniques

Tension, static compression, nanoindentation

2.1.3 Methods

Preparation of α-cellulose/spirulina slurries, 3D extrusion bioprinting (hereafter, Direct Ink Writing), lignin impregnation into bacterial cellulose (BC) sheets, BC paper fabrication, design of experiments and associated statistical methods, ternary composites fabrication, machine learning model construction method

2.2 Comprehensive Materials and Methods

2.2.1 Characterization Techniques

Scanning Electron Microscopy (SEM)

Samples were mounted with carbon tape and coated with 4 nm of platinum on a Leica EM ACE600 sputter coater (Leica Microsystems GmbH; Wetzlar, Germany). SEM imaging was performed on an Apreo VP (ThermoFisher Scientific; Waltham, MA, USA) with an accelerating voltage of 2 kV and 13 pA using a mixed detector mode of T1:T2 (BSE:SE) at 50:50.

Transmission Electron Microscopy (TEM)

Samples were prepared with an OsO_4 fixing/staining protocol prior to imaging. Small pieces of samples were immersed in deionized (DI) water and gently vortexed for 10 minutes. After allowing the samples to settle, the water was replaced with 2% OsO_4 and gently vortexed again for 10 min. This process was repeated three times with DI water to wash away excess OsO_4. The washing was then performed with increasing concentrations of ethanol in water to slowly dehydrate the samples. After washing with 50, 70, 90, and 100% ethanol the samples were washed twice with 100% acetone. The solution was then exchanged for a 50:50 mix of acetone:Mollenhauer II and gently vortexed for 4 hours. Afterwards, the samples were left to mix an additional 2 hours to allow acetone to evaporate. The samples were placed in 100% resin and left under vacuum for 2 hours to remove any remaining acetone, and were then left at 60 °C overnight to harden the resin. 80 nm sections were cut using a Leica Ultracut 6 microtome. Samples were imaged on an FEI Tecnai G2 F20 SuperTwin TEM at 200 kV with a Gatan Ultrascan CCD digital camera.

Optical Microscopy

A PantheraTEC optical microscope (Motic Instruments USA; Schertz, TX, USA) was used to study the morphology of dehydrated and hydrated spirulina powders. Transmitted light images of representative cell clusters were obtained with 20x (NA = 0.40), 50x (NA = 0.80), 100x (NA = 1.25) objectives to facilitate cell chain length and width measurements and distributions, which were all conducted using ImageJ software (http://rsb.info.nih.gov/ij/). In addition, morphological observations of the spirulina biomass were conducted on a confocal laser scanning fluorescent microscope (TSC SP8, Leica Microsystems; Wetzlar, Germany) using a 63x oil immersion objective (NA = 1.4) and imaging in white light transmission mode.

X-ray Diffraction (XRD)

Diffraction measurements were performed on a D8 Advance (Bruker; Germany) equipped with a Cu Kα source operated at 40 V and 40 A with a wavelength of 1.54184 Å. Samples were laid out on a puck with the direction of testing going along the direction of incident scattering. The degree of crystallinity was calculated using the peak height ratio, after a baseline subtraction, described by Park et al.[96] based on the following formula:

$$X_C\% = \frac{I_c - I_a}{I_c}\%,$$

where I_c is the peak intensity of the peak of interest and I_a is the height of the trough between the peak of interest and another peak.

Fourier-transform Infrared Spectroscopy (FTIR)

Samples were measured in ATR mode on a Thermo Scientific Nicolet iS10 FT-IR across a spectral range of 7800 - 350 cm^{-1} with a resolution of 2 cm^{-1} averaged over 64 runs.

Thermogravimetric Analysis (TGA)

Thermogravimetric Analysis (TGA) was performed on a Discovery TGA (TA Instruments; New Castle, DE, USA). 2 - 5 mg of sample was placed in a platinum pan and heated from room temperature to 900 °C at 20 °C/min in N_2 atmosphere.

Differential Scanning Calorimetry (DSC)

DSC experiments were performed on a TA Discovery Series DSC with ~5 mg of spirulina powder. Samples were loaded into a hermetically sealed aluminum pan with a pin-pierced hole in the lid. The samples were put through a cool-heat-cool-heat cycle to target temperatures of -70, 150, -20, 180 °C with a ramp rate of 3 °C/min and a modulation of 1 °C over a period of 1 min.

High-Performance Liquid Chromatography (HPLC)

All reactants, sugar standards, and solvents used in HPLC analyses were purchased from Millipore Sigma (Burlington, MA, USA). The chemical composition of the spirulina powder was determined based on a previously established procedure.[97] Briefly, 0.2 g of the biomass samples underwent acid hydrolysis using 3 mL of 72% (w/w) sulfuric acid for 2 hours at room temperature, in order to extract the sugars from the biomass. Subsequently, the samples were diluted with deionized water and then autoclaved at 121 °C for 1 hour, followed by filtration using sintered-glass medium porosity crucibles into a set of liquid sugar fraction and acid insoluble phenolics fraction columns. The liquid sugar fraction was analyzed using HPLC (Shimadzu HPLC system with a Rezex RHM-Monosaccharide H^+ (8%) HPLC Column 300 x 7.8 mm) for sugar content and UV/Vis spectrophotometer with absorbance reading of 205 nm (UV-1700 Shimadzu; Japan) for any acid soluble phenolics.[98] The acid insoluble phenolics fraction was measured gravimetrically after oven drying the crucibles and subtracting that mass with the mass of the crucibles before filtration.[99] Glucose, galactose, arabinose, xylose, mannose, mannitol, rhamnose, fucose, ribose, and myo-inositol were used as sugar standards for the HPLC monosaccharides analysis.

Rheological Analysis

Rheological experiments on spirulina slurries were performed on a rheometer (MCR302, Anton Paar GmbH; Graz, Austria). A small volume (1 mL) of each slurry was placed into the specimen holder at 20 °C and allowed to acclimate for 60 s. The paste samples were placed between parallel plates with an upper plate diameter of 25 mm and a gap height of 1.0 mm. Rotational measurements were performed at shear rates of 0.01–100 s^{-1} with variable time at each rate set by the instrument.

Water Contact Angle

Contact angle tests were performed using a Krüss Drop Shape Analyzer and its processing software, ADVANCE (KRÜSS Scientific; Germany), equipped with a high-resolution CCD camera. This analysis was performed in ambient conditions. Five separate DI water droplets of 4

± 1 μL were placed on each sample after mechanical testing and measured every 15 s for 2 min. The contact angle was determined according to the average angle between the baseline (sample surface) and the left and right tangent (ellipse fit) of a droplet edge.

Stylus Profilometry

Surface roughness was measured on a Dektak XTL Stylus Profilometer utilizing Vision64 analysis software (Bruker; Germany). A representative sample from each BC and LBC paper was measured in the parallel and perpendicular directions to the axis of tensile testing (measurement length: 1000 μm, duration: 10 s, force: 3 mg) and the average surface roughness (P_a) was recorded.

Viscometry

Before casting, ternary composite mixtures suspended in a 1.6 mass proportion ethanol/water solution were made in excess and a portion of the excess was used to measure the viscosity of a sample. The mixture was shaken to redistribute components and approximately 15 mL of the mixture was put into a low volume sample chamber (VOLS-1) and measured with a large spindle (VOL-SP-6.7; measurable viscosity range of 160 – 32,000 mPa•s) in a viscometer (IKA Rotavisc me-vi; IKA Works, Inc.; Wilmington, NC, USA). The shear rate, viscosity, and shear stress were recorded at variable rotational speeds of 5, 20, 30, 40, 50, 60, 70, 80, 90, and 100 rpm.

2.2.2 Mechanical Testing Techniques

Tensile Testing

Tensile testing was conducted on a mechanical test frame (AGS-X, Shimadzu Scientific Instruments; Columbia, MD, USA) using stock tensile grips with a 5 kN load cell. The tensile testing speed was 0.5%/min, adjusted accordingly for each sample depending on gauge length. All samples were desiccated for 24 hours in a 0% relative humidity chamber before being tested.

BC samples were cut directly from their respective sheets into a rectangular shape. Each end of the samples was tabbed with duct tape so that the gauge length in between the tabs measured at least 40 mm. Gauge width and thickness were nominally 4 mm and 0.1 mm, respectively, but were individually measured for each sample for properties calculations.

Chlorella-BC-stearic acid films were stamped out using a hand press (HK 2600 Hub 60 mm; Berg & Schmid GmbH, Germany) with a punch die attachment conforming to the specifications of ASTM D1708 “dogbone” tensile specimens. The specimens were weighed, duct tape tabbed at the ends of their gauge lengths, and then measured for the gauge width, thickness, and length. Samples were tested without conditioning in a desiccator beforehand.

Static Compression Testing

Compression testing was performed on 5 - 10 cubic structures of each studied group using a universal test frame (5500R, Instron; Norwood, MA, USA) equipped with a 5 kN load cell. The structures were compressed at a constant rate of 0.002 s^{-1} until failure. From the measured force (F) and displacement (d) data, we calculated the stress, σ, from $\sigma = F/A$, where A is the cross-sectional area, and strain, ε, from $\varepsilon = d/H$, where H is the initial sample height. From the stress-strain plots we obtained the compressive strength, the maximum stress before failure, the strain to break, and the stress and strain before the plateau region for samples that had a foam-like deformation. Spirulina-based structures nominally measured 1.5 cm on each side but were individually measured for each sample for properties calculations.

Nanoindentation

Nanoindentation was carried out on a Micromechanical Testing and Assembly System (FT-MTA03, FemtoTools; Zurich, Switzerland) equipped with a FT-S200000-D microforce sensing probe with a Berkovich (3-sided pyramidal) diamond tip, and a built-in optical microscope (FT-4L, 4x magnification, Optem). Displacement-controlled nanoindentation tests were performed at a constant rate of 0.1 μm/s to reach a total of 1.0 μm indentation with a maximum force of 2 mN. The load-frame compliance and area functions were calibrated on a fused quartz calibration sample (FemtoTools; Zurich, Switzerland) prior to conducting any testing on the samples of interest. Force-displacement data was analyzed using the Oliver and Pharr method to extract reduced modulus and hardness.[100]

Hardness was calculated using the equation: $H = \frac{P_{max}}{A_{plastic}}$, where P_{max} is the maximum load indentation before retracting from the sample, and $A_{plastic}$ is the residual area of plastic deformation left behind in the sample after indenting. Reduced modulus was calculated using the formula: $E_R = S \cdot \frac{1}{2\beta}\sqrt{\frac{\pi}{A_p}}$, where β is a geometrical constant (~1); $A_p(h)$ is the projected triangular indentation area and is dependent on the depth into the sample, h; S is the stiffness that is directly measured from the unloading.

To address the inherent surface roughness effects on the results, a large number of indentations (n > 20) per sample group was collected and the data was evaluated through statistical analysis.[101–103]

Data analysis for compression tests and nanoindentation

Statistical analysis of the compression and nanoindentation data was conducted by applying a two-sample Welch's t-test assuming distributions with non-equal variances. The analysis and plotting were carried out using Python (https://www.python.org/) with packages in NumPy[104], Matplotlib[105], Pandas[106], and SciPy[107].

2.2.3 Methods

Slurries preparation

Dried spirulina powder (purchased from Nuts.com) was mixed with DI water at three ratios by weight (water:spirulina): 2:1, 1.6:1, 1.2:1. A combined mass of 10 g per batch was mixed manually and then by a planetary mixer (Thinky USA; Laguna Hills, CA, USA) in 1, 2, 3, and 4 minute intervals. The slurry was allowed to cool in between mixing intervals to offset any heat generated within the slurry by mixing. Micro-crystalline α-cellulose fibers (CFs) in a powder form (Millipore Sigma; Burlington, MA, USA) were used as a filler. The water content of the spirulina/CF composite slurries was maintained at 1.2:1 weight ratio (water:solid components). The solid components were calculated such that α-cellulose fibers were 1, 5, 10, or 20 wt.% of the final dried product.

Direct Ink Writing (DIW)

The fabrication of all samples was conducted using a commercial, extrusion-based 3D bioprinter (Biobot Basic, Advanced Solutions Life Sciences; Louisville, KY, USA). The slurries were loaded in 10 mL syringe barrels with conical nozzles (1.22 mm inner diameter/16 gauge dispensing tip, McMaster-Carr; Los Angeles, CA, USA). The tip speed and acceleration were kept constant for all slurries at 12 mm/s and 10 mm/s^2. Pressure and initial stage offset were adjusted per slurry composition to maintain consistent flow without overextrusion. Set point pressure varied from 60-85 psi and stage offset varied from 1.0 - 2.0 mm depending on slurry composition. Line height and width were set at 0.70 - 0.75 mm and 0.90 - 0.95 mm, respectively, depending on slurry composition. Multilayered cubic structures with lines arranged in 0/90° orientation and a final volume of 1.5x1.5x1.5 cm^3, corresponding to 25 layers on average, were printed. The samples were allowed to set for 20 minutes, then subsequently flipped 180°, and then dried either in a desiccator, oven held at 60 °C (DVS602, Yamato Scientific; Tokyo, Japan), or lyophilizer (Freezone 2.5L, Labconco Corporation; Kansas City, MO, USA).

BC paper fabrication

BC pellicles were grown in a 30 x 60 cm tank from a symbiotic culture of bacteria and yeast starter formula (SCOBY; Joshua Tree Kombucha; USA). Growth media comprised 0.5 wt.% tea leaves (filtered out of culture media before culture began) and 5 wt.% of sucrose per liter of water. A single pellicle was harvested after 14 days of incubation, washed with water, and then soaked in a 1 M NaOH solution overnight to remove the embedded bacteria and yeast. The final BC pellicles, after extensive washing with deionized water, were divided into 10 x 10 cm^2 sheets. Each sheet was placed between water mats (New Pig Corp.; Tipton, PA, USA) and metal plates and passed through a metal roller once, before being air dried under a paperweight (~5 kg) for 7 days. Finally, each dried sheet was subjected to hot-pressing following a design of experiments explained below.

The processing steps are summarized in **Fig. 2.1**.

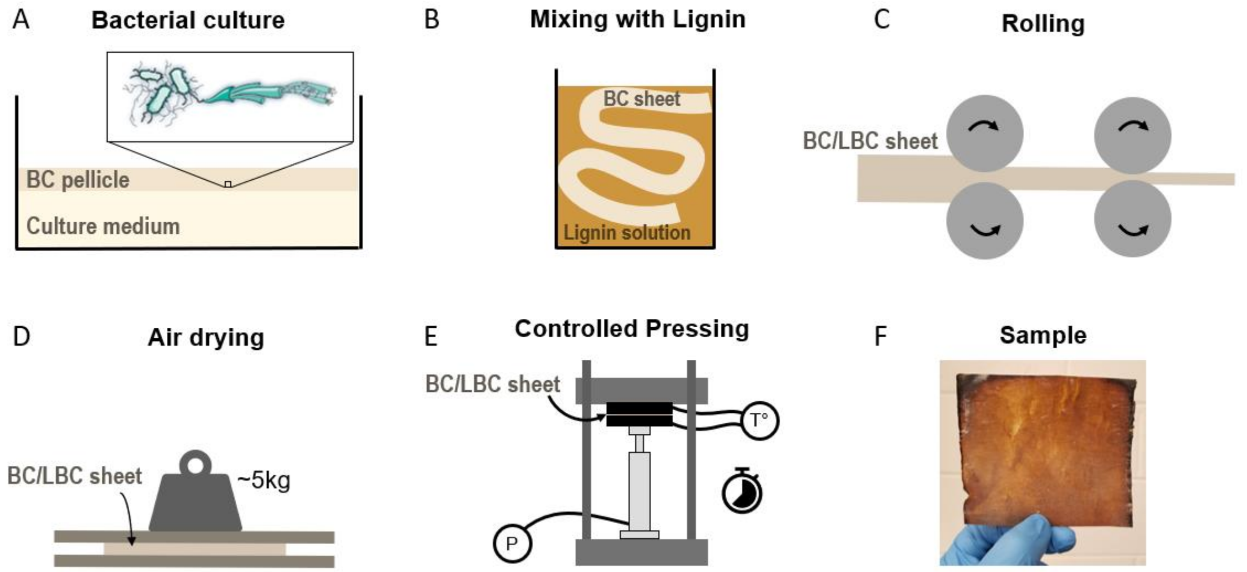


Figure 2.1: Schematic of bacterial cellulose (BC) sheet processing. (a) BC is cultured in a tank as a pellicle and then cut into smaller sheets. (b) Optional processing step of impregnating BC with lignin via stirring in a dissolved lignin solution. (c) Sheets are rolled out in one direction to remove water and help align fibrils. (d) Sheets are dried under weight until paper-like in thickness and touch. (e) Cold-pressing followed by hot-pressing of sheets under DOE conditions of a specified temperature, pressure, and time. (f) Representative image of a resultant composite sample after processing.

Lignin impregnation into bacterial cellulose sheets

A 0.1 g/mL lignin solution was prepared by dissolving sulfonated alkali lignin (TCI America; Portland, OR, USA) in water, stirring under ambient temperature. BC sheets were placed in the solution and stirred at 120 rpm for 1 hour before being removed. The resulting lignocellulose sheets (LBC) were placed on a water mat for 3 minutes on each side to remove excess lignin solution, and subsequently rolled and processed as described in the *BC paper fabrication* methods. After the papers were pressed into sheets (as described in the following *Design of Experiments and statistical methods* section), a Klason analysis was performed, measuring 15.1 ± 1.9 wt.% (n = 6) of lignin in the composites.

Design of Experiments and statistical methods

BC and LBC sheets were processed under the same design of experiments (DOE) conditions. All samples were initially cold-pressed at 40 °C for 5 minutes at 5 MPa before being hot-pressed. Hot-pressing conditions varied in temperature (120 - 160 °C), pressure (5 - 15 MPa), and time under pressure (10 - 30 minutes). Parameter combinations were assigned in a Latin squares design (See Table 2.1). A longhand notation is used here where “temperature (°C)/pressure (MPa)/time (min.)” is used to refer to the factor-levels of the DOE sample (e.g. “120/5/10” refers to DOE 1, which shares the same factor-levels).

Property-parameter dependence was determined using the Minitab software (Minitab, LLC; State College, PA). A general factorial regression model was fitted to the data for main effects and two-way interactions analyses. For statistically significant main effects ($p < 0.05$) a post-hoc Games-Howell analysis at 95% confidence was conducted with one-way ANOVA for further pairwise difference comparisons. Lack-of-fit error analysis was used to determine whether a main effects model was appropriate for the data: a significant value ($p < 0.05$) indicated that the main effects model did not fit the data.

Table 2.1: DOE design: factor-level combinations

T (^oC)	5 MPa	10 MPa	15 MPa
120	**DOE 1** / 10’	**DOE 2** / 20’	**DOE 3** / 30’
140	**DOE 4** / 20’	**DOE 5** / 30’	**DOE 6** / 10’
160	**DOE 7** / 30’	**DOE 8** / 10’	**DOE 9** / 20’

Key:
(e.g.) DOE 1 = “120/5/10”

Ternary (chlorella-BC-stearic acid) composites fabrication

Dry stearic acid (reagent grade – 95%; Sigma-Aldrich) and chlorella powder (*Chlorella vulgaris*; Nuts.com; Cranford, NJ, USA), were used as-is (without pre-processing) in the measurement of specified compositions. A dry mass calculation was used to determine the amount of BC to add to the composite as wet mass hydrogel (“hydrated BC”) as the third component of the hydrogel. The batch-to-batch dry mass of cellulose can vary in cultured bacterial cellulose, necessitating the measurement of the dry mass of each batch. Representative portions from BC pellicles in a batch were cut out, weighed, and then dried in a 65 °C oven. Dried portions were weighed and then an average dry mass per hydrated mass was calculated. The resulting quotient was taken to be the dry mass percent of cellulose in a hydrated portion of bacterial cellulose.

Once a composition to be made was determined, a wet mass of BC pellicle corresponding to the dry mass composition, dry chlorella powder, and dry stearic acid flakes were measured out. The chlorella and stearic acid were combined and mixed in a planetary mixer (SpeedMixer DAC 330-100 PRO, FlackTek; Landrum, SC, USA) in the following automated protocol: 1000 rotations per minute (rpm) for 10 seconds, 4 seconds rest, 3500 rpm for 90 seconds, 10 seconds rest, and then 3500 rpm for 90 seconds. The heat generated from the mixing was sufficient to cause melting of the stearic acid into the chlorella, indicated by the absence of distinct stearic acid flakes and the hardening of the resulting chlorella mass.

The BC pellicle portion and combined stearic acid and chlorella mass were put into a 450 mL blender cup. A mass of ethanol corresponding to 1.6 times the mass of the water in the hydrated BC was measured and added into the cup. The mixture was blended in an immersion blender (1400 peak wattage Ninja Professional Plus Blender DUO, SharkNinja Operating LLC; Needham, Massachusetts, USA) using the following protocol: 15 seconds of continuous blending (at the maximum speed preset by the blender) followed by removal of the cup from the blender and shaking by hand to distribute the components to the bottom of the cup. This protocol was sequentially repeated for 10 sets for a total of 150 seconds of blending.

The resulting mixture was cast onto a flat tray lined with Teflon and then flattened with a spatula, when necessary. This process was necessary when the amount of solvent used did not exceed the solvent holding capacity of the dry mass components, allowing the formation of a workable slurry. When the amount of solvent exceeded the holding capacity, there was clear separation between the liquid solvent and dry mass after casting and settling.

The casted mixture was set in a non-air-circulating oven set at 30 °C until the solvent evaporated off and the film was dry and easily removed from the tray surface.

Machine learning model construction method

Dry mass compositions of ternary composites were used as features in a machine learning algorithm, from which a model relating the composition to the magnitude of mechanical properties was constructed. Training data consisting of dry mass compositions and mechanical properties (labels/target variables: Young's modulus, tensile strength, elongation at max stress, and toughness) were loaded into the algorithm and trained on an exact GP model from the Python *GPyTorch* library or a Gaussian Process Regressor model from the *scikit-learn* library. The likelihood model used was the Gaussian likelihood. The kernel was set to RBF (radial basis function) and an initial length scale of 0.55 with the "fmin_l_bfgs_b" optimizer. The model and likelihood were trained using the Adam algorithm optimizer with a learning rate of 0.01 for 250 iterations. The model produced from the training data was validated on test data, which was designated to be approximately 10% of the total data. The acquisition function used an expected improvement algorithm. In this algorithm, the expected improvement is calculated as the sum of the likelihood that a particular point is higher than the current best value from the training set, and the uncertainty at a given point. The objective function was set as a default maximizer of a single mechanical property.

3. Spirulina cell-based hierarchical nanocomposites through Direct Ink Writing

3.1 Materials and Methods

See **Chapter 2** for comprehensive details on the methods used. In brief, SEM was used to image cross-sections of dried composites, nanoindentation was used to probe surface-level properties of composites, compression was used to test macroscopic mechanical performance of printed composites, OM was used to image and analyze wetted spirulina particles, TEM was used to visualize intercellular bridging after processing the composites, HPLC was used to elucidate the cell wall components of spirulina, FTIR was used to confirm HPLC results and the presence of peptidoglycans, and DSC was used to understand the gelatinization behavior of spirulina slurries after heating.

3.2 Results and Discussion

3.2.1 Pure spirulina morphology

Spirulina cells naturally aggregate as interconnected chain units. **Fig. 3.1a** shows a representative optical microscopy image of hydrated spirulina displaying its distinctive cell chains. The width of a chain is dictated by the width of the individual cells that comprise it, while the length of a chain is determined by the number of cells within a chain. The length and width size distributions based on measurements in optical microscopy images are summarized in **Fig. 3.1b-c**. Hydrated cell widths are fairly consistent among cells at 2-10 µm, while lengths vary between 10-50 µm. SEM images of the dehydrated spirulina powder (**Fig. 3.1.1d-e**) show a collapsed chain morphology and confirm that the biomass consists of intact cells and their chained structures.

Spirulina cells develop four-layered cell walls[108,109], with the innermost layers, L-I and L-II, comprised of structural glucan and peptidoglycan layers, respectively, thus providing structural stability to the cell wall.[110] TEM images (**Fig. 3.1f-g**) revealed that the cell wall and protoplasmic features of the cells remained largely intact within the biomass used to create the slurries. The cell wall thickness measured 30.3 ± 1.2 nm, which varied between 27.3 ± 1.9 nm in the surrounding parts of the cells, and 33.3 ± 1.4 nm in the cross-wall regions (junction between two neighboring cells of the same chain, marked with arrows in **Fig. 3.1g**).

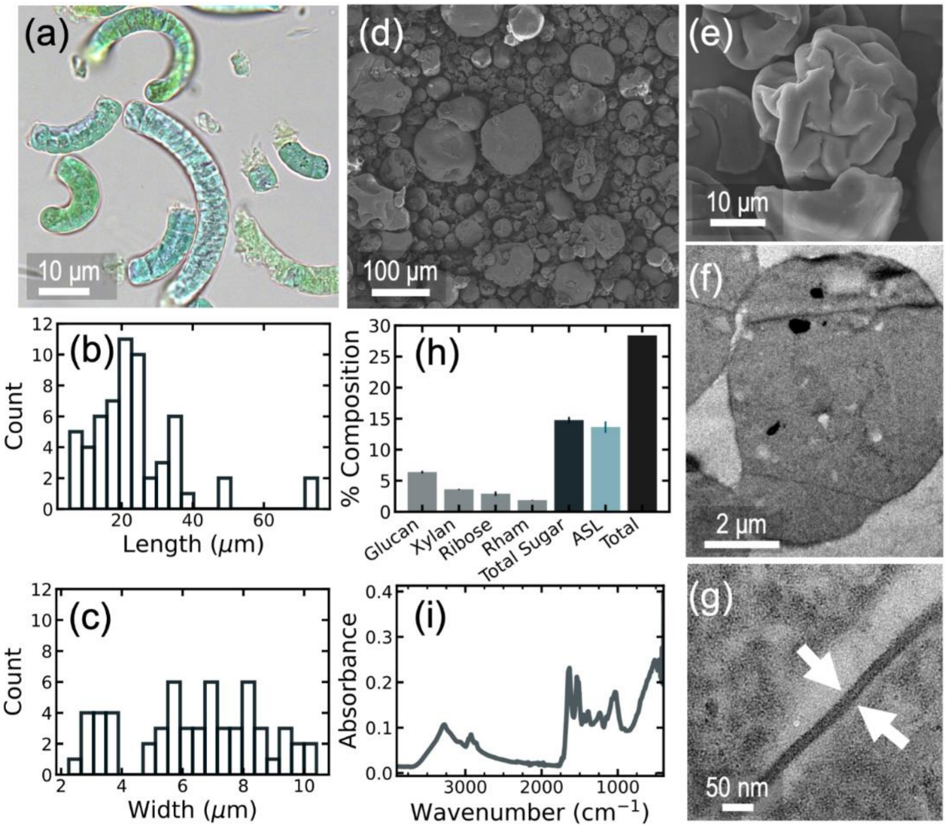


Figure 3.1: (a) Optical microscopy image of hydrated spirulina powder. (b-c) Distributions of length and width of spirulina chains. (d-e) SEM and (f-g) TEM images of dried spirulina powder. Arrows indicate the cell wall between two cells of the same chain. (h) Sugar analysis of spirulina powder obtained from HPLC. ASL stands for acid-soluble lignin. (i) FTIR spectrum of spirulina powder.

The monosaccharidic composition of the spirulina biomass was characterized by a combination of acid hydrolysis, HPLC, and UV-Vis analyses, and the results are summarized in **Fig. 1.1h**. The analyses revealed the presence of glucose, xylose, ribose, rhamnose, and acid soluble phenolics in spirulina, which are consistent with literature [108,111], though in different relative concentrations owing to variations in species and growth conditions. The detected glucose residues may be components of sulphated lipoglucan polymers [112] or peptidoglycans, which are amino sugars (*N*-acetylglucosamine + *N*-acetylmuramic acid) cross-linked with oligopeptide chains,[113] both of which have been reported in literature for spirulina cells. Thus, the monomers identified in our chemical analyses, combined with literature, suggest the presence of peptidoglycans or lipoglucans in the biomass used in this work.

FTIR spectroscopy on the spirulina powder **Fig. 3.1i** allowed further insights into the molecular makeup of spirulina biomass. Peaks assigned to carbohydrates include 1035 cm^{-1} for C-O and C-C stretching[44,114–117], 1394 cm^{-1} for carboxylate group vibrations[44,114–117], 2855 and 2873 cm^{-1} for C-H stretching in aldehyde groups[115,118], and the 2855 - 2960 cm^{-1} for the symmetric and asymmetric stretching of C-H bonds in CH_2 and CH_3 groups[44,110,116,118–120]. The hydrogen bonding region showed peaks at 3280 and 3066 cm^{-1} that arise from carbohydrates as well as amino acids[44,110,114,115,120]. Protein contributions were detected through the peak at 1538 cm^{-1} which is attributed to N-H scissor bending in the amide II[44,110,114,116,119]. Lastly, the peak at 1640 cm^{-1} was attributed to amide I C=O stretching and NH_2 scissor bending in amino acids, particularly in the peptidoglycan L-II layer[44,114–116,118,119,121]. Therefore, combining the chemical analyses and FTIR results, we confirmed the presence of peptidoglycans in the spirulina biomass.

3.2.2 Pure spirulina direct ink writing

We aimed to create fully biobased extrudable slurries for room temperature 3D-printing, employing a natural spirulina powder. First, we evaluated pure spirulina slurries without any additives by varying the water to biomass weight ratio from 2:1, to 1.6:1, and 1.2:1. We performed rheology tests (**Fig. 3.2**) and measured the viscosity of each of the slurries over a range of shear rates (0.01 to 100 s^{-1}) to understand the stresses exerted on the slurries during extrusion printing. All of the samples showed a continuous shear thinning behavior that in similar biomass-based slurries has been attributed to continuous polymer chain alignment through the shearing process.[122] Specifically, as shear rate increases, the interactions between the polymers in the slurry are disrupted and molecular chains can slide and move more easily.[123,124]

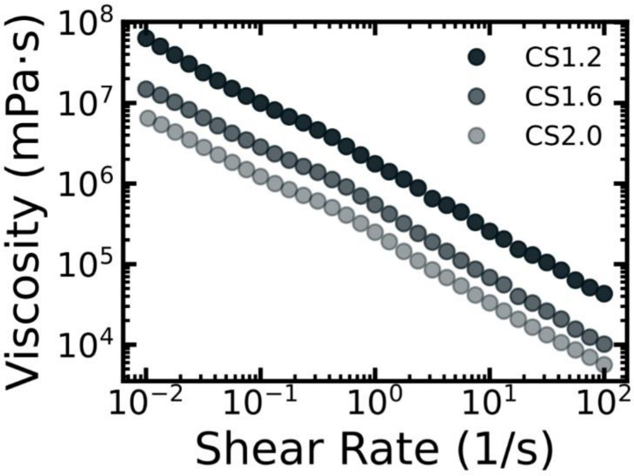


Figure 3.2: Rheology of pure spirulina slurries with different water content.

There was a progressive increase in viscosity, which spans more than one order of magnitude across our slurries, with decreasing water content. We confirmed that the slurry with the lowest water concentration (1.2:1; 55 wt.%) and highest viscosity led to the most consistent extrusion process with regard to immediate post-print shape retention. Slurries with higher water content had reduced shape retention after the extrusion process. Therefore, 55 wt.% water content was optimal for creating spirulina-based, extrudable slurries.

We subsequently 3D-printed multilayered cubic structures of volume 1.5x1.5x1.5 cm^3 using the slurry, which corresponded to 25 layers made of parallel lines alternating between 0/90°

orientations per layer. Following the printing process, we subjected the cubic samples to one of three different drying methods and evaluated the effects of each of these post-processing steps on the shape retention, micro-morphology, and mechanical properties of the printed samples. We chose drying in a desiccator, oven, and lyophilizer (freeze-dryer) in order to obtain distinctly different micro-morphologies and densities that we hypothesized would lead to different macroscopic mechanical properties. Desiccator-drying was selected to produce a densified morphology. Oven-drying was used to generate a densified and plasticized morphology. Finally, freeze-drying was applied to create highly porous microstructures. The ability to create three distinctive micro-morphologies effectively allowed tuning of the density of the structures without altering their geometry and provided control of the final mechanical properties.

To study the water-induced plasticization, or gelatinization, of spirulina we conducted DSC measurements on the pure powder (**Fig. 3.3**). The recorded endothermic transition showed a peak at 61 °C, which suggested that incubating the hydrated samples at that temperature after printing induces gelatinization in the biomass. Upon cooling and subsequent heating the same transition was not observed, demonstrating that the plasticization was maintained and not reversed upon cooling. Thus, a temperature of 60 °C was selected to process the oven-dried samples.

In the pure spirulina slurries, drying in the desiccator and oven resulted in cracks that catastrophically deformed the structures in every attempt (**Fig. 3.4c,e**), proving the need for a filler material to mitigate the drying induced stresses. Freeze-drying allowed the wet print to maintain its shape after drying, demonstrating an option to use a pure spirulina slurry (i.e. without any additives or binders) as a 3D-printing material for extrusion-based printing at room temperature.

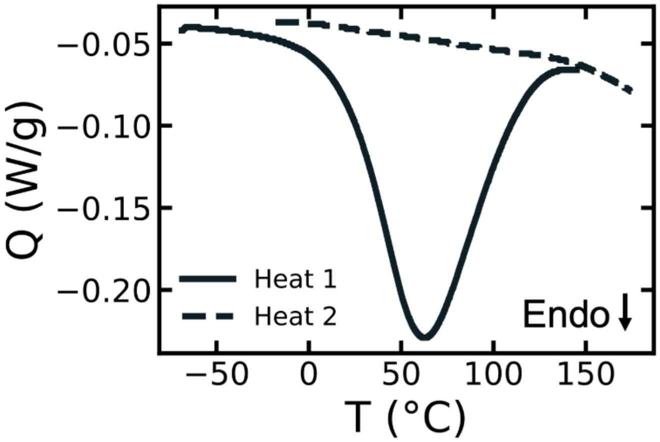


Figure 3.3: DSC thermograms of spirulina powder in subsequent cycles (Heat 1 vs Heat 2). The first heat showed a broad endothermic peak from 0 - 150 °C, with a maximum at 60.9 °C. After cooling, the immediate subsequent heat showed no endothermic peak in this range.

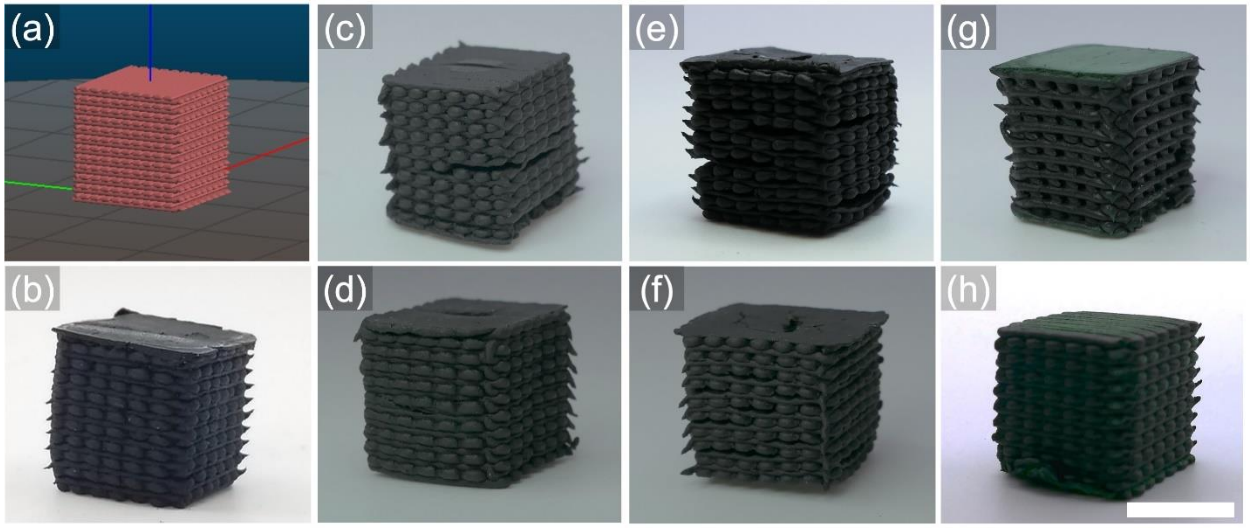


Figure 3.4: 3D-printed spirulina-based structures. (a) Model rendering, (b) pure spirulina structure after printing, before drying. Printed structures after desiccator-drying: (c) spirulina, (d) 5 wt.% CF composite. Printed structures after oven-drying: (e) spirulina, (f) 5 wt.% CF composite. Printed structures after freeze-drying: (g) spirulina, (h) 5 wt.% CF composite. Scale bar is 1 cm.

The gravimetrically measured densities of the pure spirulina structures dried with each of the three methods were 0.85 ± 0.01, 0.65 ± 0.01, and 0.54 ± 0.01 g/cm^3, for the desiccated, oven-dried, and freeze-dried samples, respectively. Despite fracturing during drying, desiccated prints imaged in SEM show the continuous, densified, self-bonded matrix that is formed throughout the bulk of the print (**Fig. 3.5**). Intact spirulina cells which maintain their shape and are immersed in the continuous matrix can be distinguished (**Fig. 3.5b-i**). The continuous matrix is itself a homogenized phase of dissociated, ruptured cells (**Fig. 3.6**). More specifically, the protoplasmic cell contents as well as cell wall fragments, which appear as "nanoflakes" (marked by arrows in **Fig. 3.6**), form a self-bonded, amorphous, continuous phase. In the following discussion, we refer to this miscible blend phase as the *amorphous matrix*. The spirulina cells are fully bonded within their surrounding amorphous matrix but form a distinct phase, giving rise to a heterogeneous blend that is referred to as *bulk blend* hereafter.

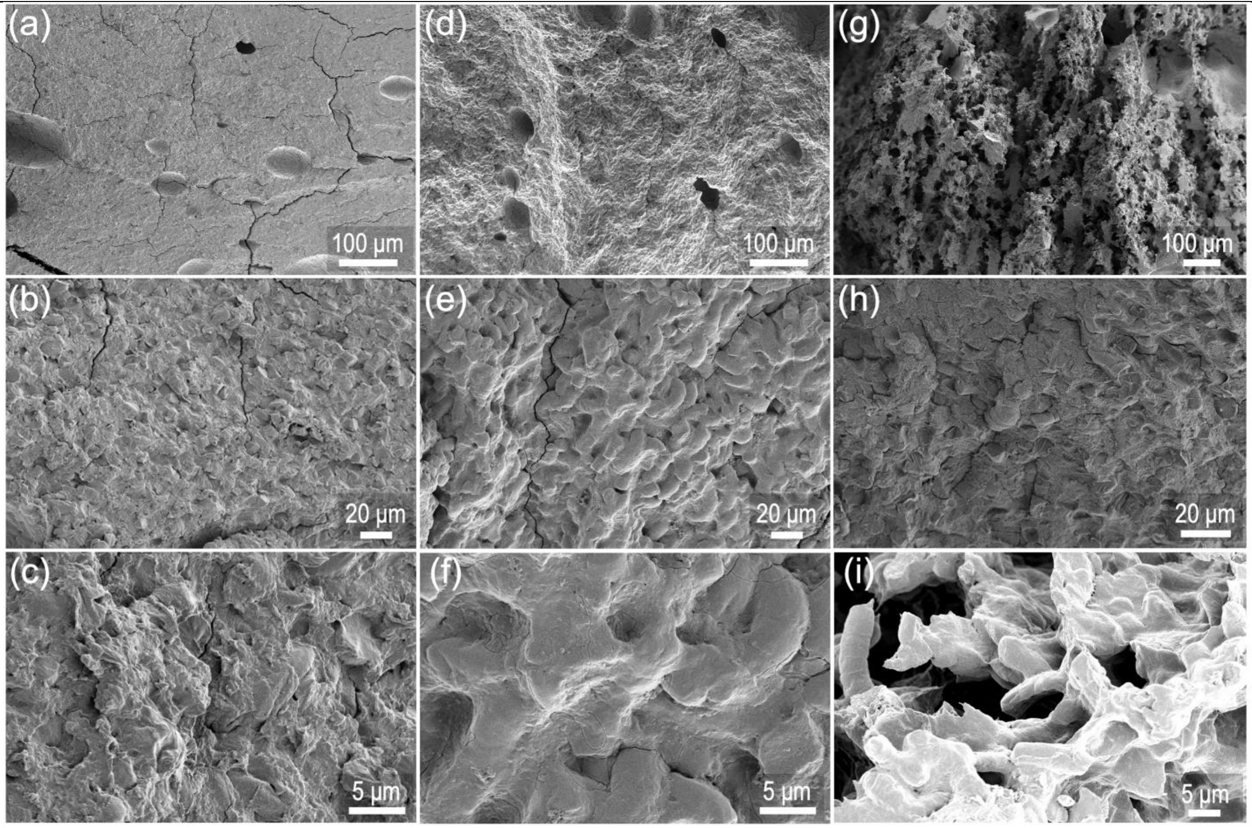


Figure 3.5: SEM images of pure spirulina dried with each method. (a-c) Desiccator, (d-f) oven, (g-i) freeze-dried materials.

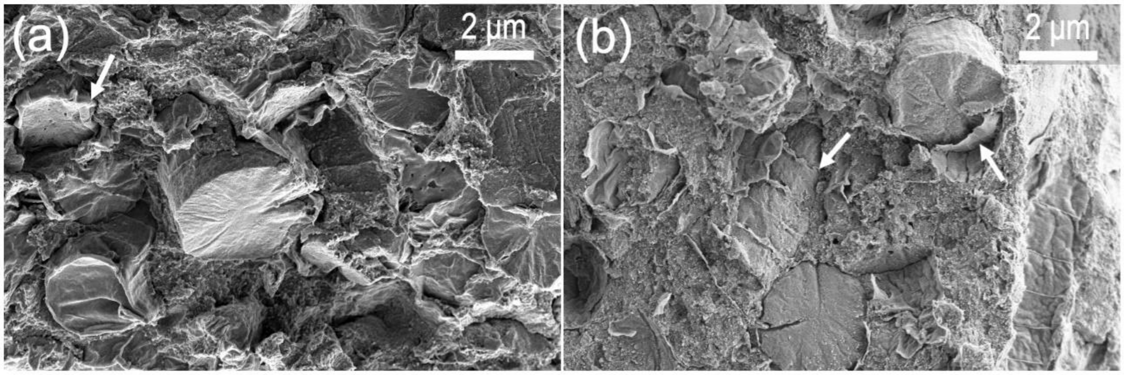


Figure 3.6: SEM images of the morphological features of the cell/matrix blend phase. Views of (a) 1 wt.% and (b) 5 wt.% CF composites after drying in the desiccator and oven, respectively. Arrows pointing to cell wall flakes.

Fig. 3.5e-f demonstrate that oven-drying the printed samples at 60 °C facilitates matrix plasticization, as the characteristic "fused" morphology of gelatinized biomass[125] was observed. Additionally, the drying temperature is low enough to not destroy the cell structures, as intact cell imprints were seen in the SEM images. Another difference between the oven- and desiccator-dried materials is the amount of intact cells observed in the matrix. The images of the oven-dried samples revealed a greater amount of cells (42 ± 4 cells per 1000 μm^2), while desiccated samples appeared to have fewer whole cells (8 ± 1 cells per 1000 μm^2) and a greater proportion of partially or entirely ruptured cells. The homogeneous, amorphous matrix was more prevalent in the desiccated micro-structure as well.

In the freeze-dried samples, two distinct regions were formed: a dominant foam region and a small amount of a bulk, densified region. The bulk region had the same feature of cells immersed in an amorphous matrix as the bulk blends seen in the desiccated samples. In the foam region of the freeze-dried samples, intact cells were still fully embedded in the continuous amorphous matrix, which formed a homogeneous coating around the cells, similar to the bulk blends. The high porosity of the foam at the micro- and meso-scales differentiate it from the structures obtained from the other two drying methods.

By imaging the fracture surfaces of the samples, the cracks formed within the bulk regions were observed (**Fig. 3.5a, d, g**). The oven-dried and desiccated samples showed that a large network of micro- and nano-cracks formed within the bulk blend. Desiccated surfaces (**Fig. 3.5a**) predominately had cracks of micron-sized width (micro-cracks), substantially larger than those found on oven-dried sample surfaces (**Fig. 3.5d**) which showed more nano-sized cracks (nano-cracks). The bulk areas of the freeze-dried samples also showed more nano-cracks. In addition to the micro- and nano-cracks, large pores with diameters of 100s of microns were observed in the desiccated samples, while smaller diameters of pores were noted in the oven-dried samples. In the bulk blends, the cracks were present in either the amorphous matrix or at the interface between cells and the amorphous matrix. Crazing features were also found in the amorphous phases of the bulk blends, contributing to the failure modes of our materials. In the freeze-dried samples, the foam regions dominated deformation behavior and offered energy dissipation mechanisms.

3.2.3 Composites of spirulina and α-cellulose: printability, structure, and mechanical properties

Introducing CFs in the spirulina slurry was studied as a strategy to mitigate the detrimental effects of drying and improve the mechanical properties of the printed materials. We studied the effects of incorporating CF fibers in printed spirulina structures on the morphology and mechanical properties of the resulting composites subjected to each of the three post-processing methods. We prepared spirulina/CF composite slurries with 1, 5, 10, and 20 wt.% CF, keeping the water content across all slurries constant at 55 wt.%, the optimal content in the pure spirulina slurries.

Rheological tests of the composite slurries revealed the same shear thinning behavior as seen in the pure spirulina slurries. **Fig. 3.4** shows representative images of dried 5 wt.% CF composites, in comparison to pure spirulina samples, demonstrating the beneficial effects of the filler on shape retention for each of the drying methods.

SEM images reveal that the CFs were well incorporated in the matrix, as we observed continuity between the primary matrix and the filler at all prepared concentrations (**Fig. 3.6**). The presence of the filler caused some changes in, but did not substantially alter, the micro-morphology of the spirulina matrix. The composites still consisted of CFs and a blend matrix that is itself a heterogeneous mixture of cells in an amorphous phase, similar to the control samples. However, the ratio of cells/amorphous matrix, which was higher in the oven-dried samples for pure spirulina slurries compared to the desiccated samples (42 vs 8 cells per 1000 μm^2), was different in the composites. At the studied concentrations, an almost equal ratio of cells/amorphous matrix was observed for the oven-dried and desiccated composites. For the 1 wt.%, for example, we measured 17 ± 2 vs 20 ± 1 cells per 1000 μm^2 for oven and desiccated samples, respectively. Additionally, the oven-dried samples showed less prominent plasticization compared to pure spirulina. In the controls, all the bulk matrices showed a "fused", plasticized micro-morphology (**Fig. 3.5e,f**). Yet, in the composites, plasticized areas were present (**Fig. 3.7b,c**), but did not entirely comprise the bulk of those samples. This difference suggests that the presence of CFs may inhibit the water uptake and swelling of the polymers in the biomass. Because CFs are extremely hydrophilic, they would compete with the rest of the biomass for water uptake, thus reducing the biomass plasticization in the composites.

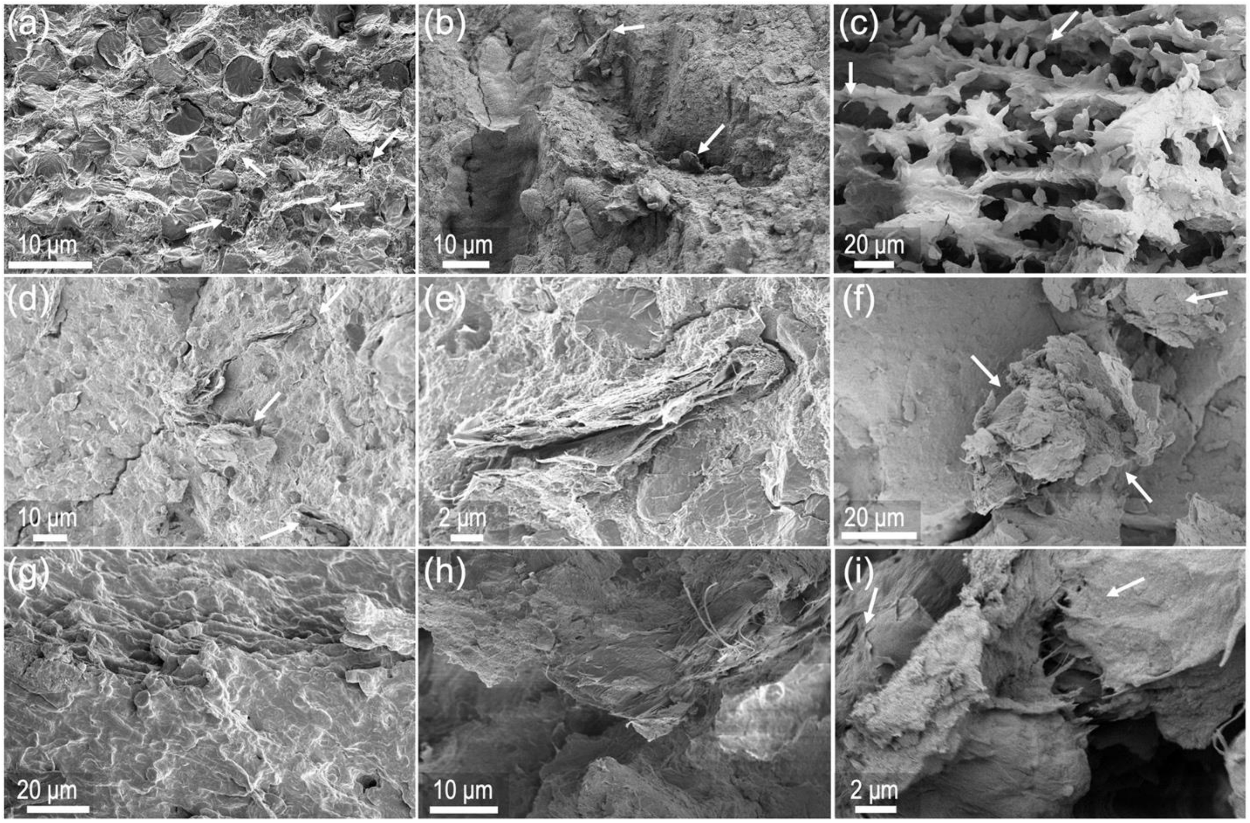


Figure 3.7: SEM images of the composite structures. (a) 1 wt.% CF desiccated, (b) 5 wt.% CF oven-dried, (c) 10 wt.% CF freeze-dried, (d-e) 10 wt.% CF oven-dried, (f) 10 wt.% CF desiccated, (g-h) 20 wt.% CF oven-dried, and (i) 20 wt.% CF freeze-dried. Arrows point to CFs.

At concentrations below 10 wt.%, the CFs were well-dispersed and formed "exfoliated" sheet-like bundles (*flakes*). We noticed no debonding or pull-out for these exfoliated flakes. On the contrary, the filler was well-bonded with the matrix, with excellent wetting of the CF fibers as evidenced by the uniform amorphous matrix (**Fig. 3.7**). At 10 wt.%, changes in the clustering and distribution of CFs were noted for two different drying methods. While in the oven-dried samples CFs maintained a fine distribution and formed small flake-like bundles (**Fig. 3.7d,e**), large cluster formations were found in the desiccated samples (**Fig. 3.7f**). At the highest CF loading (20 wt.%), the largest clusters seen across any of the samples were formed, with entire cluster pull-out from the fractured surfaces (**Fig. 3.7h**), as well as debonding and failure within the cluster with both drying methods.

The CF distribution in the freeze-dried samples was different from the other composites. In the freeze-dried composites, the fibers formed a coating around the blend, both in the cell-rich and the amorphous-rich areas (**Fig. 3.7c,i**). The fractured surfaces revealed nanofibrils from the CF coating being pulled out or still connecting the matrix together. Even at the highest filler loading, the CFs were not found to create large clusters, but were spread as a coating over the matrix with bonding and continuity between the two components in all cases.

All the freeze-dried composites had lower densities than the pure spirulina samples (Table 3.1). Additionally, the volume of the composite freeze-dried prints was 9 - 26% higher compared to the pure spirulina, following no particular trend with increasing filler content. The presence of the fibers allowed the prints to shrink less than the control and dry evenly in the oven and desiccator without failing, unlike in the pure spirulina prints, highlighting the mitigating effects of CFs on the stress distribution during drying and on strengthening the spirulina matrix so that it sustains higher stresses before failure.

Table 3.1: Density and mechanical properties of pure spirulina and composite structures tested in compression

Drying method	CF concentration (wt.%)	Density (g/cm^3)	Post-drying volume (g/cm^3)	Compressive strength (MPa)	Strain to failure (%)
Freeze-dried	0	0.54[a] ± 0.01	2.94 ± 0.25	0.8 ± 0.1	12.1 ± 2.2
	1	0.52 ± 0.02	3.32 ± 0.56	0.8 ± 0.2	16.4 ± 3.9
	5	0.52 ± 0.01	3.51 ± 0.26	1.6 ± 0.3	24.5 ± 3.7
	10	0.50 ± 0.01	3.25 ± 0.23	2.7 ± 0.5	27.3 ± 2.3
	20	0.51 ± 0.02	3.83 ± 0.65	6.6[a] ± 1.5	47.3[b] ± 10.1
Oven-dried	1	0.68 ± 0.02	2.38 ± 0.47	0.7 ± 0.1	23.1 ± 5.3
	5	0.69 ± 0.03	2.52 ± 0.17	2.1 ± 0.4	23.3 ± 2.6
	10	0.66 ± 0.04	2.61 ± 0.24	3.7 ± 1.1	31.9 ± 2.2
	20	0.64 ± 0.02	2.61 ± 0.20	10.3 ± 0.7	37.7 ± 3.1
Desiccator-dried	1	1.01 ± 0.02	1.70 ± 0.19	3.2 ± 0.7	21.2 ± 4.7
	5	0.99 ± 0.07	2.11 ± 0.38	4.5 ± 3.1	21.3 ± 8.3
	10	0.87 ± 0.07	1.87 ± 0.10	6.6 ± 1.1	31.3 ± 3.2
	20	0.91 ± 0.06	2.22 ± 0.38	16.4[c] ± 4.7	29.2[d] ± 5.0

Key: a – Plateau stress; b – Strain at the beginning of plateau region; c – Stress at end of experiment, not at failure; d – Strain at end of experiment, not at failure.

We subjected all the composite structures and the freeze-dried control samples to uniaxial compression tests to evaluate the impact of CFs in the (global) mechanical properties of spirulina. In **Fig. 3.8** we present characteristic stress-strain plots from compression tests across all the material concentrations and drying methods, as well as the distributions of the compressive strength values, respectively. For the freeze-dried controls we noticed a linear increase in stress with strain followed by a point with maximum stress before fracture (compressive strength of the material), which is followed by progressive cracking leading to fracture at 12.1 ± 2.2% strains. In those samples, the measured compressive strength was 0.8 ± 0.1 MPa. The freeze-dried composites showed a progressive enhancement of the compressive strength for CF concentrations higher than 1 wt.%. There was no difference in strength between the (pure) control and the 1 wt.% CF composite, but there was a statistically significant difference in the strain at failure, with the composite able to sustain strains of 16.4 ± 3.9% before failing. This increase in strain to failure indicates that the CFs facilitate a greater extent of polymer chain sliding during compression within the biomass matrix, or they utilize the supplied energy to move, possibly debundle, and spread themselves. CFs facilitating polymer chain sliding is more likely, and can be supported by effects reported in pure cellulose materials.[19,126] Due to the large amount of surface hydroxyl groups in CFs, as has been shown in literature through both experiment and simulation, when cellulose chains are subjected to load there is a cascading hydrogen bond breaking and re-forming, which allows a higher extensibility before failure and ultimately improves toughness.[19,126] This effect is more pronounced as the CF diameter is reduced.[19] Thus, it is possible that CFs form hydrogen bonds with their surrounding polymer matrix, and when compressed, the CFs or matrix polymer chains experience this breaking and reformation of hydrogen bonding which ultimately improves the strain before fracture, which is what we experimentally observed. To summarize, the introduction of 1 wt.% CF in freeze-dried samples led to an improvement in the deformation before fracture, as well as a slight reduction of density and shrinkage, yet no improvement in strength.

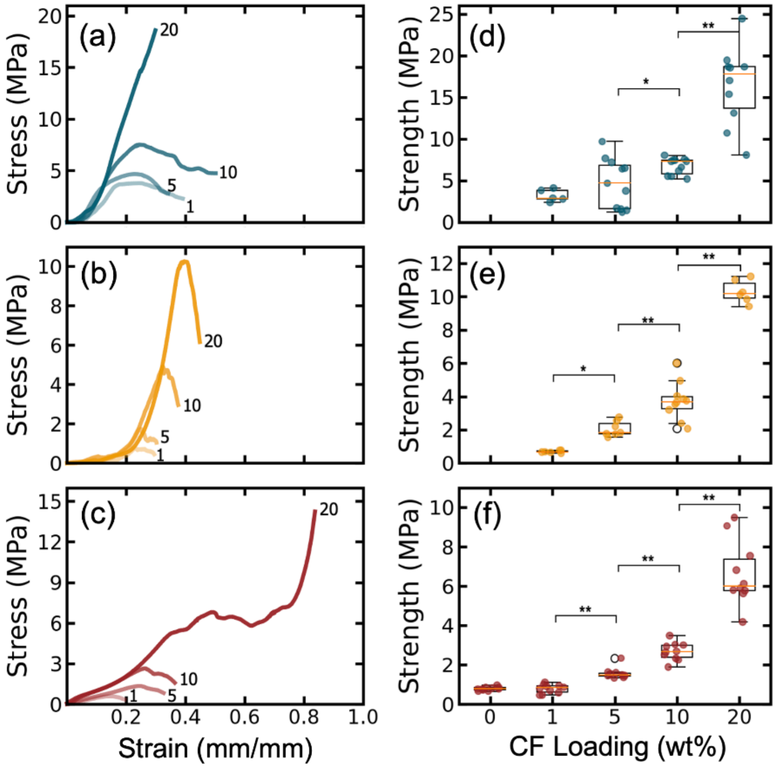


Figure 3.8: Stress-strain plots for (a) desiccated, (b) oven-dried, and (c) freeze-dried structures. Compressive strength versus filler concentration for (d) desiccated, (e) oven-dried, and (f) freeze-dried structures. The 20 wt.% freeze-dried sample strength corresponds to plateau stress, and the 20 wt.% desiccated sample strength values are the stress at end of experiment, not at failure. * $0.1 < p < 0.01$, ** $p < 0.01$

For higher concentrations of CFs, we measured significantly higher compressive strength values as well as higher strains at failure, indicating an overall toughening of the composite with higher filler loadings. For 5 and 10 wt.% freeze-dried composites the stress-strain curves showed the same behavior as the control samples - a linear loading followed by a point of maximum stress after which the material fractures. The average compressive strength values of the 5 and 10 wt.% composites were 1.9 and 3.2 times higher than the pure spirulina’s, while the strains to break were double. At the highest CF loading, 20 wt.%, a major change in the deformation behavior was observed. The stress-strain curves of the 20 wt.% composites revealed a predominantly foam-like deformation, with a linear-elastic region followed by a plateau and finally a densification region (**Fig. 3.8c**). We note that these composites have the highest volume (lowest shrinkage) among all freeze-dried samples. The average plateau stress of the 20 wt.% composite was 8.2 times higher than the compressive strength of the control sample, and the deformation at the beginning of the plateau region was almost 4 times higher than the total strain to break of the controls.

The progressive increase in stress we saw up to 20 wt.% suggests that these composites had not reached a percolation threshold regarding their mechanical reinforcement as no strength saturation was observed.[127] All the freeze-dried composites had slightly lower densities than the pure spirulina samples (statistically significant differences ranging from 3.7% to 7.4%). The significant increases in strength and strain to break with increasing filler concentrations, the excellent adhesion and distribution of the filler observed through SEM, and the reduction in shrinkage for all composites confirm that the cellulose fibers act as an effective reinforcement for the spirulina matrix.

The effects of CFs in the oven-dried and desiccated samples are even more pronounced than the freeze-dried, as samples without any filler content were lacking structural stability, and deformed and cracked severely upon drying, making macroscopic testing impossible. On the other hand, all composites with CFs, even at the lowest concentration of 1 wt.% were able to be dried and tested in compression. Since pure spirulina controls dried with these two methods are structurally not stable, we could not quantify the effects of CF on the mechanical properties of the prints. Thus, the following comparisons are among composite prints.

Oven-dried samples had a substantially lower density compared to the desiccated samples, with average values ranging between 0.63 - 0.69 g/cm^3 and 0.85 - 1.00 g/cm^3, respectively. The volumes of the oven-dried samples were 25% higher than the desiccated samples', while both oven- and desiccator-dried samples had 35-45% lower volume than the freeze-dried samples. We conclude that freeze-drying leads to the minimum shrinkage and lowest density in composite samples, followed by oven drying which leads to intermediate shrinkage and density, and desiccation resulting in the highest shrinkage and highest density.

The consistently lower strengths of the oven-dried composites, compared to their desiccated counterparts with the same filler concentration were justified by the lower density of the oven-dried samples. Plasticization leads to deformation behavior differences of samples dried in the oven compared with those dried in the other two methods. The stress-strain curves of oven-dried composites reveal a slow initial increase in stress, at an almost constant stress for the first 10 - 20% of the deformation, before the bulk response is probed, followed by a brittle fracture. In freeze-dried and desiccated samples, there was no such distinct region before the bulk behavior was probed. We note that the oven-dried samples were fully in contact with the compression plates during that initial period, which implies that those plasticized composites dissipate the supplied energy possibly through polymer chain rearrangement or sliding before the collective, densified behavior dominates the deformation response.

In composites dried either in the oven or desiccator, we measured a progressive increase in the strength and strain to break with increasing filler content for both data sets up till 20 wt.%, where the deformation behavior changes. With the CF concentration increasing from 1 to 5 to 10 wt.%, the average compressive strength values increased 2.8 and 4.8 times in the oven-dried, and 1.3 and 1.8 times in the desiccated samples, respectively. For the 10 wt.% desiccated composites, there is a bimodal distribution in the deformation behavior, with a small subset of the samples (3 out of 11 tested) exhibiting a foam-like deformation, with a linear region followed by a plateau and a final densification region, while the majority of the samples (8 out of 11) followed the same brittle fracture as the other composites. There was no difference in densities for those two populations within the 10 wt.% group. However, SEM revealed that in that particular group (**Fig. 3.7d-f**), CFs were distributed as large clusters or dispersed flakes. In comparison, only clusters were found at higher loadings while dispersed flakes were seen at lower loadings. Thus, this bimodal state of CFs could justify the bimodal deformation behavior seen in the compression tests of the 10 wt.% desiccated composites, with some samples demonstrating behavior stemming from a dispersed-CF-dominant state, and some from a clustered-CF-dominant state.

The highest density, desiccated samples at 20 wt.% showed a uniquely elastomer-like deformation behavior. That is, its stress-strain profile had a hyperelastic region of monotonic increase in stress with strain without a yielding behavior or brittle fracture. Oven-dried 20 wt.% composites had the same deformation behavior as oven-dried samples at all other CF loadings, with an average compressive strength over 13.7 times higher than the 1% oven-dried composite and an almost 37% strain to break, the highest among the oven-dried composites group.

The compressive strength values of the composites at a density range of 0.5 - 1.0 g/cm^3, are comparable to those reported for 3D-printed amylopectin/cotton fiber composites,[50] which utilized gelatinization at higher temperatures (80 - 120 °C) to create a cohesive amorphous matrix. When evaluated against synthetic plastics commonly used in extrusion-based 3D-printing, our biobased composites have comparable compressive strength and higher strain to break. Printed structures of ABS (density = 1.04 g/cm^3) have a compressive strength of about 20-40 MPa and strain to break of 15%.[128–130] Comparatively, PLA has a higher density (1.24 g/cm^3), higher strength (40 - 45 MPa), and a much lower strain to break (5 - 10%).[128,129] Therefore, the performance of the biobased composites presented in this work is comparable to plastics commonly used for 3D-printing, as well as pectin/cotton fiber composites of similar density.

The effects of CFs on the compressive strength of the composites were more pronounced upon matrix plasticization, as evidenced by oven-drying showing the largest comparative compressive strength, followed by freeze-drying and then desiccator-drying. Excellent CF dispersion and adhesion to the matrix enabled a gradual increase in compressive strength and strain to failure with loading, effectively toughening the material at all concentrations in all drying methods. At the highest filler contents, 10 and 20 wt.%, we noticed that deformation behavior changes among the freeze-dried and desiccated composites, which obtain foam-like and elastomer-like behaviors depending on the drying method, while at all other concentrations, the composites failed brittlely.

3.3 Conclusions

In this chapter, a series of extrudable polymer composites consisting entirely of natural biopolymers from spirulina biomass and cellulose fibers were prepared and used to create 3D-printed multilayered structures. The effects of altering the morphology and density of the printed structures in order to tune their mechanical properties were discussed. The density of the structures can be altered by drying them in a desiccator, oven, or lyophilizer. Freeze-drying leads to the lowest shrinkage and lowest density across all drying methods, followed by oven-drying, which plasticizes the matrix, and finally desiccation, which results in the most densified structures.

The introduction of cellulose fibers substantially enhances the printability and shape-retention ability of the 3D-printed composites, in addition to enhancing the mechanical properties of the produced structures. Without the filler, the structures are unable to retain their designed shape, drying in an oven or desiccator with substantial fractures. However, the inclusion of cellulose fibers at concentrations as low as 1 wt.% allows the multilayered structures to retain their shape, regardless of the drying method.

The excellent dispersion of cellulose fibers and their adhesion to the spirulina matrix leads to a progressive increase in the compressive strength for all the composite structures with increasing filler concentration. The highest filler concentration (20 wt.%) improves the compressive strength over 8.2 times and the strain to break almost 4 times compared to the pure spirulina freeze-dried samples. Even though the composites with the highest density and filler content, desiccator-dried 10 and 20 wt.% CFs composites, have the highest observed filler agglomeration, their mechanical properties still follow the same increasing trends with filler content as the rest of the composites.

This work demonstrates that purely biobased composites can be adapted to extrusion-based 3D-printing without requiring the use of any non-degradable additives or curing agents, and upon minimal post-processing, robust meso-scale multi-layered structures can be produced. Finally, we show that the mechanical properties of the produced structures can be tuned either from the post-processing method or the relative concentration of the biopolymer components.

4. Effects of lignin on bacterial cellulose nanocomposite materials

4.1 Materials and Methods

See **Chapter 2** for comprehensive details on the methods used. In brief, we used SEM to image the surface layers and fracture cross-sections of composites, XRD to measure the degree of crystallinity in the cellulose components, and TGA to understand how lignin impacted cellulose degradation characteristics. Tensile testing was performed to understand the mechanical properties (Young's modulus, strength, strain to break, and toughness) of composite strips, water contact angle measurements were conducted to understand the change in water absorption behavior of composites after processing with and without lignin, and stylus profilometry was used to inform whether the water repellent behavior of the composites could be explained by the processing or the physical topography of the composites.

4.2 Results and Discussion

4.2.1 Effects of lignin in bacterial cellulose

Prior work has indicated that lignin may improve binding between cellulose fibrils in a cellulose matrix[16,24,25,131]. In **Fig. 4.1a** we present the results of tensile testing of all the BC and LBC papers (prepared under the conditions reported in **Table 2.1**, in **Section 2.2.3 Methods**). Overall, the incorporation of ~15 wt.% lignin leads to LBC composites with significantly improved elastic modulus, and marginally improved strength, at the expense of toughness. There are 66% and 9% increases in Young's modulus (E; 9.9 to 16.4 GPa) and ultimate tensile strength (UTS; 128.0 to 139.6 MPa), respectively, and a 58% decrease in toughness (Tgh; 2.48 to 1.04 MJ/m^3) in the lignin-containing composites. The average density for all LBC samples was 1.38 g/cm^3 compared to 1.04 g/cm^3 for BC samples.

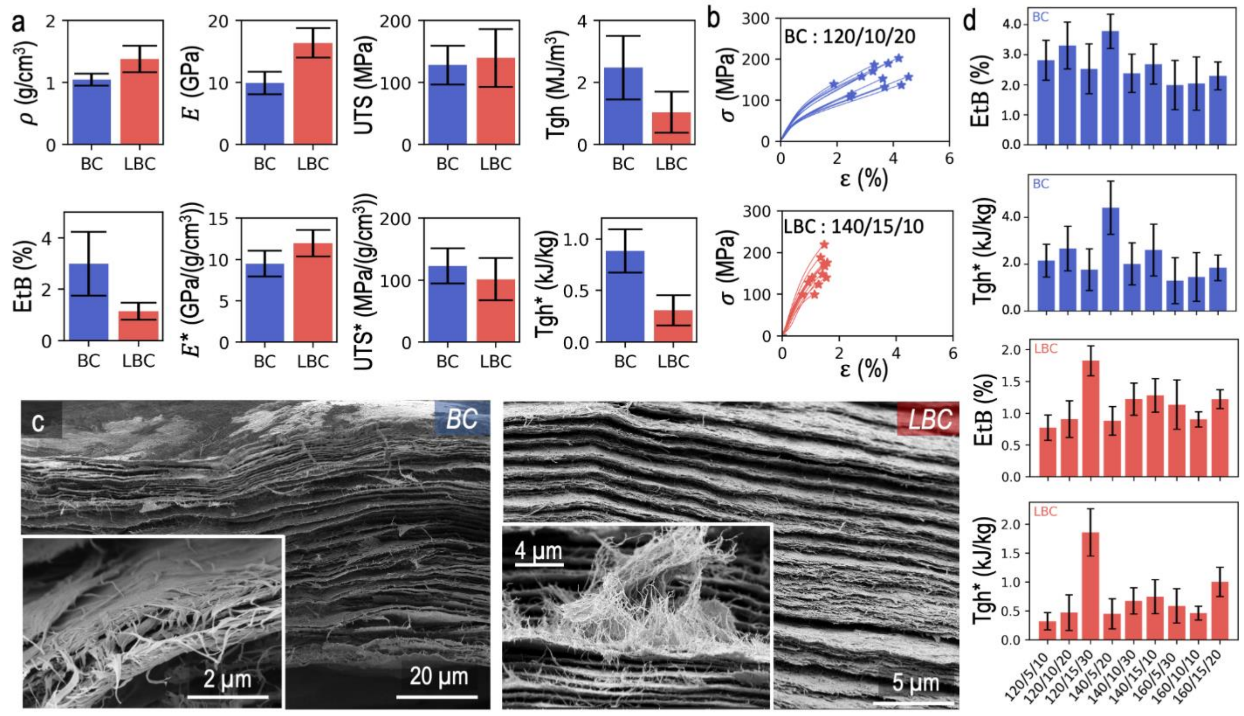


Figure 4.1: Mechanical properties of BC (blue) and LBC (red) papers. Starred values indicate properties normalized by sample density. (a) Comparative average properties across all DOEs. *ρ* is density, EtB is elongation to break, *E* is Young's modulus, UTS is ultimate tensile strength, and Tgh is toughness. The starred properties are normalized to sample density. (b) Representative stress-strain curves for a BC and LBC sample. (c) SEM images depicting the layered structure of the BC and LBC papers at the fracture surface. Insets show individual cellulose fibrils as the primary load-bearing element in these systems. The BC paper is pressed at 160/5/30, and the LBC at 140/5/20. (d) Comparative bar graphs of each DOE condition in BC and LBC for elongation to break and specific toughness. Due to the brittle nature of the samples, EtB has proportional behavior to Tgh*.

Starred values, indicating specific properties, were normalized against individual sample densities and then averaged. Comparing normalized values between the BC and LBC samples confirms the observed positive influence of lignin in the Young's modulus and its reduction of the toughness of the produced papers. Analyzing the specific UTS (UTS*), we observe a marginal strength reduction in the LBC composites compared to BC, though the standard deviations in both cases (UTS and UTS*) overlap.

SEM analysis shows the morphology of pressed BC papers as layered sheets of cellulose fibers (**Fig. 4.1c**). The addition of lignin does not fundamentally change this structure at the level of lignin incorporation studied here, as seen in **Fig. 4.1d**. Cross-sectional views of the fracture surface indicate the individual cellulose fibrils as the load-bearing element in both BC and LBC papers, with no distinct lignin phase appearing at the fracture surface of the composite (**Fig. 4.1c** inset). The increased stiffness seen in LBC samples suggests that the lignin is well-incorporated into the matrix and mediates binding between cellulose fibrils. However, the decrease in LBC UTS* suggests that lignin acts as an origin for tensile failure, possibly due to the formation of weak, intralayer aggregates. Such aggregates may prevent cracks from reaching the individual cellulose fibrils, thus promoting stress concentration in the weaker lignin-rich areas. Furthermore, lignin does not contribute ductility to BC, as indicated by the brittle failure at UTS in representative stress-strain curves (**Fig. 4.1b**). Elongation to break (EtB, equivalent to EtB*) showed a modest difference between BC and LBC: 3% vs. 1%. This decrease could be explained by the hindrance of cellulose chain sliding caused by hardened lignin interfaces between fibrils.

4.2.2 DOE analysis of the effects of pressing temperature, pressure, and time on the produced papers

To maximize the mechanical properties and minimize property trade-offs from the introduction of lignin, we investigated the effect of temperature, pressure, and time during the hot-pressing step on the tensile properties of our materials. We hypothesized that these three factors could alter the extent of densification, as well as lignin-cellulose interactions, as prior literature reports demonstrated that hot-pressing temperature alone (around the established transition temperature of lignin, i.e. 80 - 150 °C) allows tuning of the crosslinking and interactions between cellulose and lignin.[24,132] Each factor was divided into low, medium, and high values (levels) for a design of experiments (DOE, for value ranges see **Section 2.2.3** and Table 2.1). Mechanical properties for each processing condition are reported in **Fig. 4.2**.

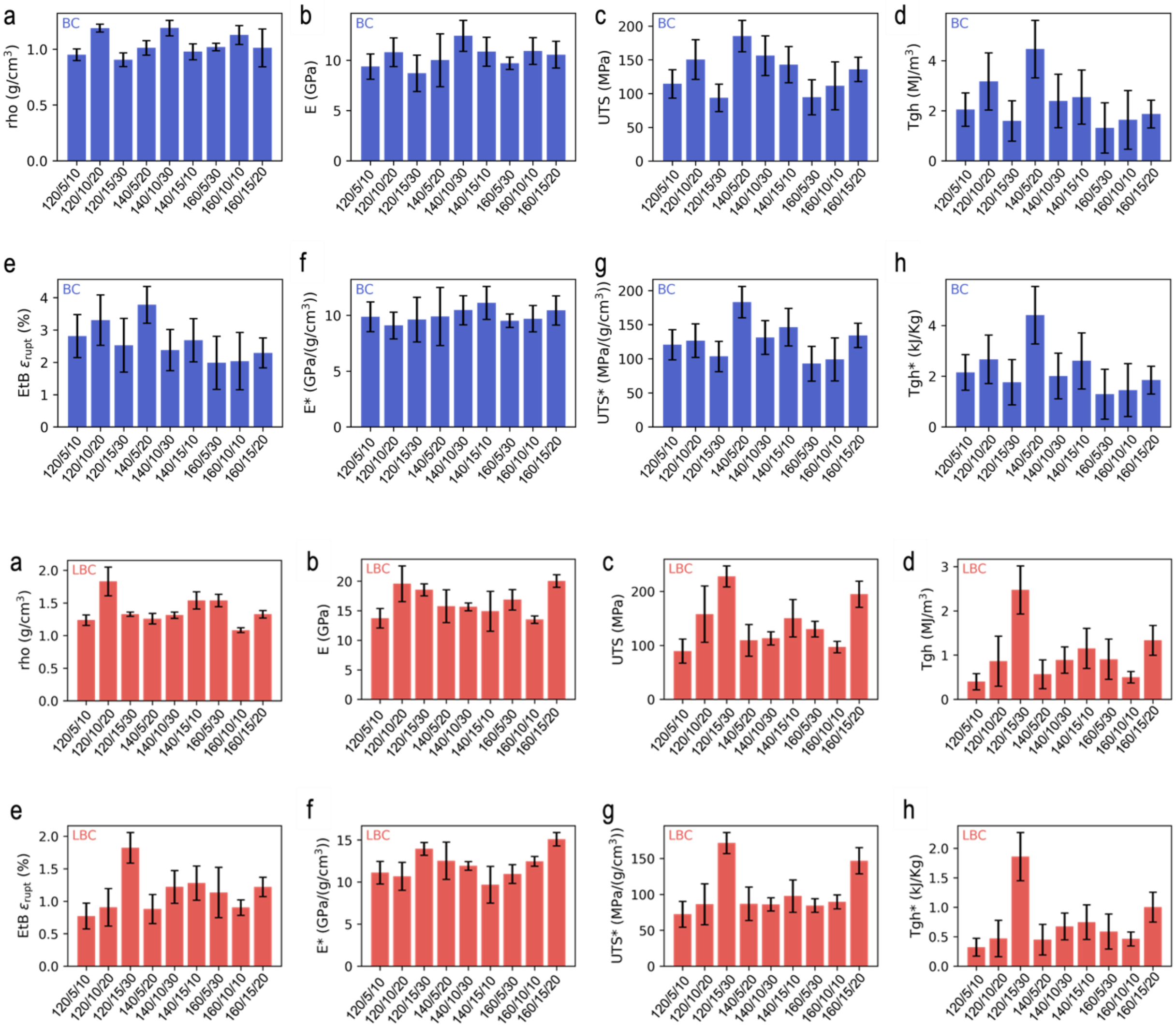


Figure 4.2: All mechanical properties for each processing condition for BC (blue, top panel) and LBC (red, bottom panel). “rho” refers to density. All other properties are the same as reported in **Fig. 4.1**. Bottom labels indicate separate samples with conditions corresponding to that in Table 2.1.

A Latin squares design minimizes the number of necessary combinations of experiments required to infer parameter-property relationships, but all main effects are confounded with interaction effects as a result.[133] Thus, main effects must be analyzed with potential interactions to be able to draw conclusions. Main effects plots based on a factorial regression model were generated but are not presented here – main effects that were significant and resulted in a maximum or minimum for properties are summarized in Table 3.2. The regression model assumed no interaction parameters; i.e., all response variation could be explained by a linear combination of each factor-level pair. The one-way ANOVA lack-of-fit errors indicated that this model was appropriately applied for all four specific mechanical responses (E*, UTS*, EtB, Tgh*) for BC samples and EtB for LBC ($p > 0.05$). However, $p < 0.05$ for LBC E*, UTS*, and Tgh* suggested that interactions play a prominent role in these property responses, or that the responses would have a best fit by nonlinear models. *p*-values from the model ANOVA were used to evaluate whether any of the means at each level were significantly different from each other. Significantly different means were further analyzed using a post-hoc Games-Howell test to distinguish the main effect(s) in a sample.

Table 3.2: Significant extrema main effects, where the blue cells are maximum, and the red cells are minimum.

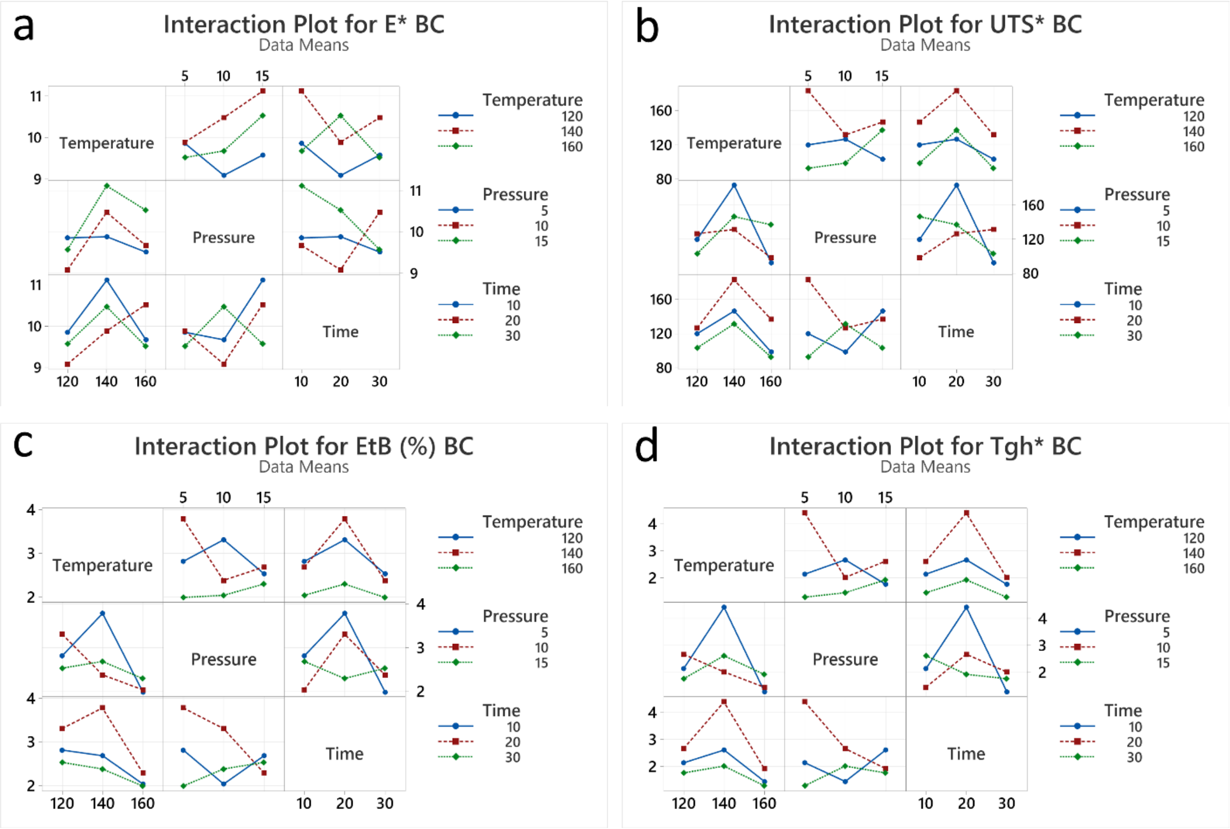
Main effects @ $p < 0.05$		**BC**	**LBC**		**BC**	**LBC**		**BC**	**LBC**		**BC**	**LBC**
T	**E***	🟥 120 °C	120 °C	**UTS***	🟥 120 °C	120 °C	**EtB**	🟦 120 °C	120 °C	**Tgh***	120 °C	120 °C
		🟦 140 °C	🟥 140 °C		🟦 140 °C	🟥 140 °C		🟦 140 °C	140 °C		🟦 140 °C	140 °C
		160 °C	🟦 160 °C		🟥 160 °C	🟦 160 °C		🟥 160 °C	160 °C		🟥 160 °C	160 °C
P		5 MPa	🟥 5 MPa		5 MPa	🟥 5 MPa		5 MPa	🟥 5 MPa		5 MPa	🟥 5 MPa
		10 MPa	🟥 10 MPa		10 MPa	🟥 10 MPa		10 MPa	🟥 10 MPa		10 MPa	🟥 10 MPa

		15 MPa	15 MPa		15 MPa	15 MPa		15 MPa	15 MPa		15 MPa	15 MPa
t		10 min	10 min		10 min	10 min		10 min	10 min		10 min	10 min
		20 min	20 min		20 min	20 min		20 min	20 min		20 min	20 min
		30 min	30 min		30 min	30 min		30 min	30 min		30 min	30 min

Key: T = temperature, P = pressure, t = time under pressure

Overall, there were no main effects found to consistently alter the mechanical properties of both BC and LBC samples. In BC, time and temperature were shown to have a significant effect on the final mechanical properties, while pressure did not. The prominent effects of temperature and time are anticipated, as they would directly affect the amount of bound water in the BC papers.[22,134] Molecular dynamics simulations have shown that the inter-fibrillar bound water molecules will spontaneously leave their confinement at temperatures around 160 °C,[134] which upon cooling would lead to differences in the degree of inter-fibrillar interactions within the BC papers. In LBC, all three factors had a level showing a main effect. For temperature, 140 °C resulted in maxima for E*, UTS*, and Tgh* in BC while the same temperature resulted in minima for E* and UTS* in LBC. In contrast, 160 °C maximized properties in LBC while minimizing them in BC. For time, 20 and 30 minutes resulted in the best and worst UTS*, EtB*, and Tgh* in BC, respectively. Pressing times of 30 minutes and 10 minutes resulted in the best and worst of the same properties for LBC, while 20 minutes maximized only E*.

Conclusions from these main effects can only be drawn in the context of the interactions between our selected processing conditions, which are not accounted for in the main effects plots. **Fig. 4.3** depicts the qualitative two-way interaction plots to elucidate the coupled effects of our processing conditions in the final mechanical properties of our materials. For example, no pressure-time dependence could be discerned for BC. Pressure had no main effect in BC, so the values of the interaction plots should be proportional to time. Observing the time-pressure interaction plots, in general, 20 minutes maximizes and 30 minutes minimizes the UTS*, EtB, and Tgh* as anticipated from the main effects results stated earlier.



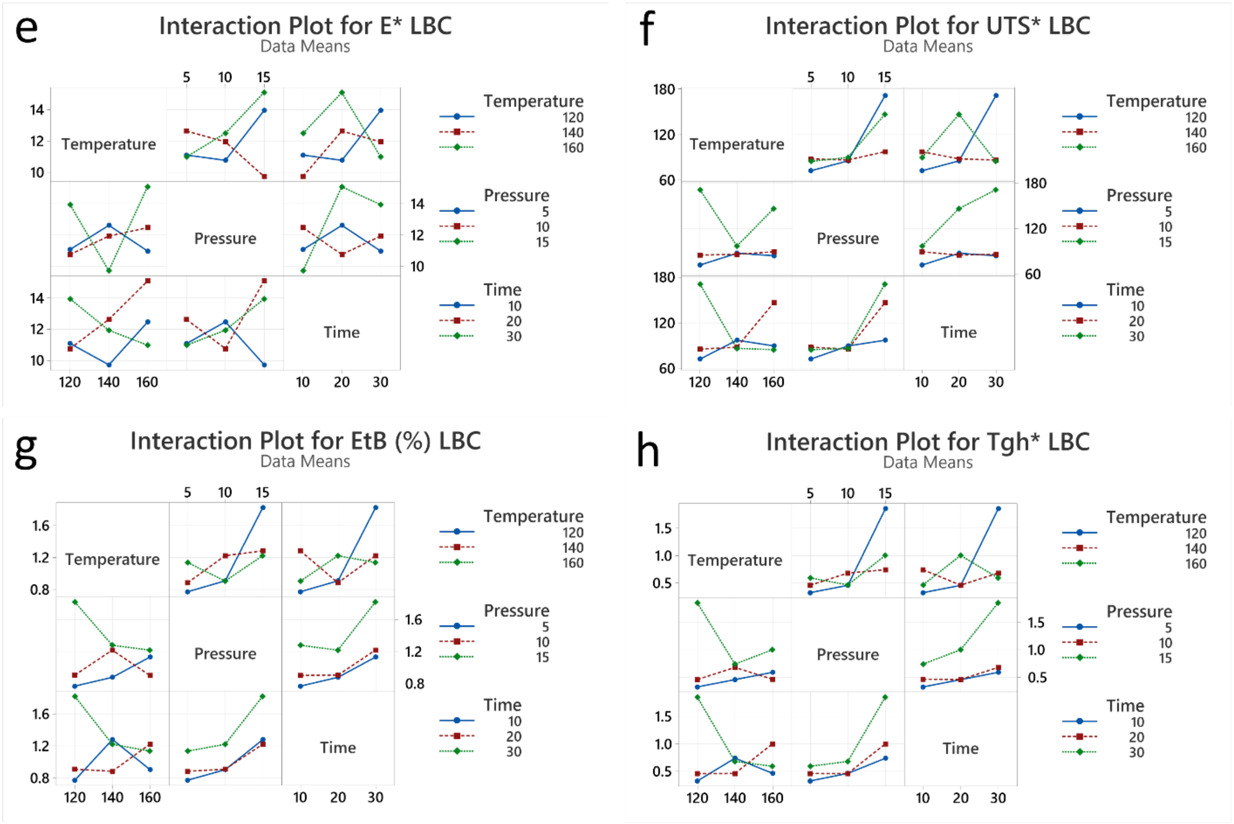
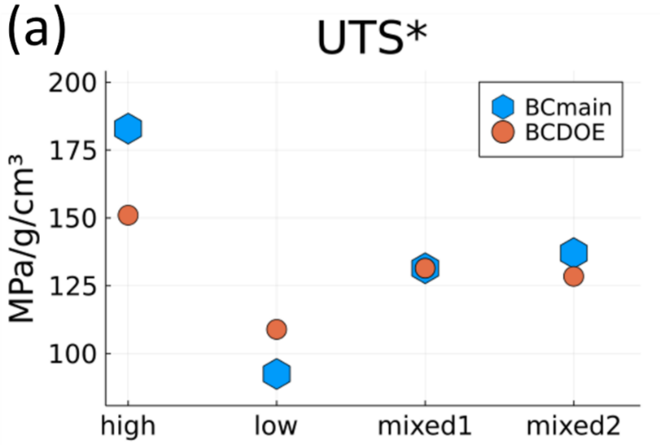


Figure 4.3: Two-way interaction plots for (a) BC: E*, (b) BC: UTS*, (c) BC: EtB, (d) BC: Tgh*, (e) LBC: E*, (f) LBC: UTS*, (g) LBC: EtB, and (h) LBC: Tgh*.

Lack-of-fit errors relate to the fit of a linear model. A linear model can inform whether a positive effect on performance is supported or attenuated by another factor. If a factor-level increases the performance of a material in a linear system, then we can be reasonably certain that that factor-level alone is the reason for the improvement. To validate the lack-of-fit error analysis, we can take combinations of the extrema of performances from the main effects and compare against the DOE that corresponds to their two-way interactions. For a linear main effects model, the performances of combined main effects are equivalent to that of interaction pairs, since interactions effects are small or nonexistent. That is, the average of two main effect properties should be similar to the same two-way interaction's properties. In contrast, if interactions are strong and a main effects model does not fit, the theoretical combination of two main effects may not be the same as the actual properties of a two-way interaction. The properties for which main effects of BC exist and can be compared against its two-way interactions are UTS*, EtB, Tgh*. Specifically, we compare a combination of temperature and time since there are no main effects for pressure. The factor-levels that produced the highest, lowest, and a combination of the highest and lowest main effects were plotted in **Fig. 4.4**. From these values, we can tell if the linear model is correct based on how close the DOE and main effect combinations are to each other. In all cases, the properties of the mixed values lie between that of the extrema. Furthermore, the values of the DOE and predicted main effects are close, with an average percent difference of 10% (UTS*), 24% (Tgh*), and 14% (EtB). These small differences suggest that interaction effects, if present, are small in the BC papers.



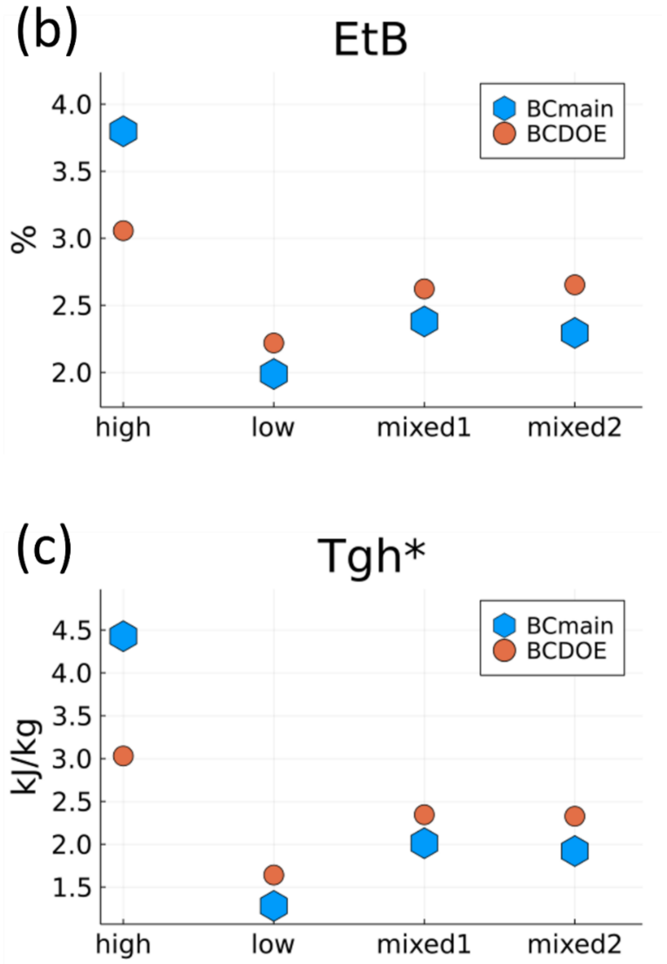


Figure 4.4: Temperature-time factor-level combination for BC main effects vs. DOE comparison. Where BCmain: theoretical combination of main effect, BCDOE: realized combination of two factors, "high": the property-maximizing combination, "low": the property-minimizing combination, "mixed1"/"mixed2": A combination consisting of one factor-level that maximizes and another that minimizes properties.

Due to the possibility of three-way interactions in LBC papers, paired combinations of all three factors were analyzed (**Fig. 4.5**). In these cases, the mixed main effects all fell between the high and low values. However, the mixed DOE properties did not consistently fall between these values, and their positive or negative deviations could not be predicted by the values of the main effect combinations. Compared to the BC UTS* temperature-time interaction, the equivalent for LBC showed a larger average percent difference between the DOE and main effects values (BC

10% vs. LBC 18%), again supporting that the linear main effects model does not apply for LBC because interactions are significant. Temperature-time and temperature-pressure appeared to have the largest interactions in LBC papers, based on the relative deviations of the measured mechanical properties to the theoretical main effects combination properties.

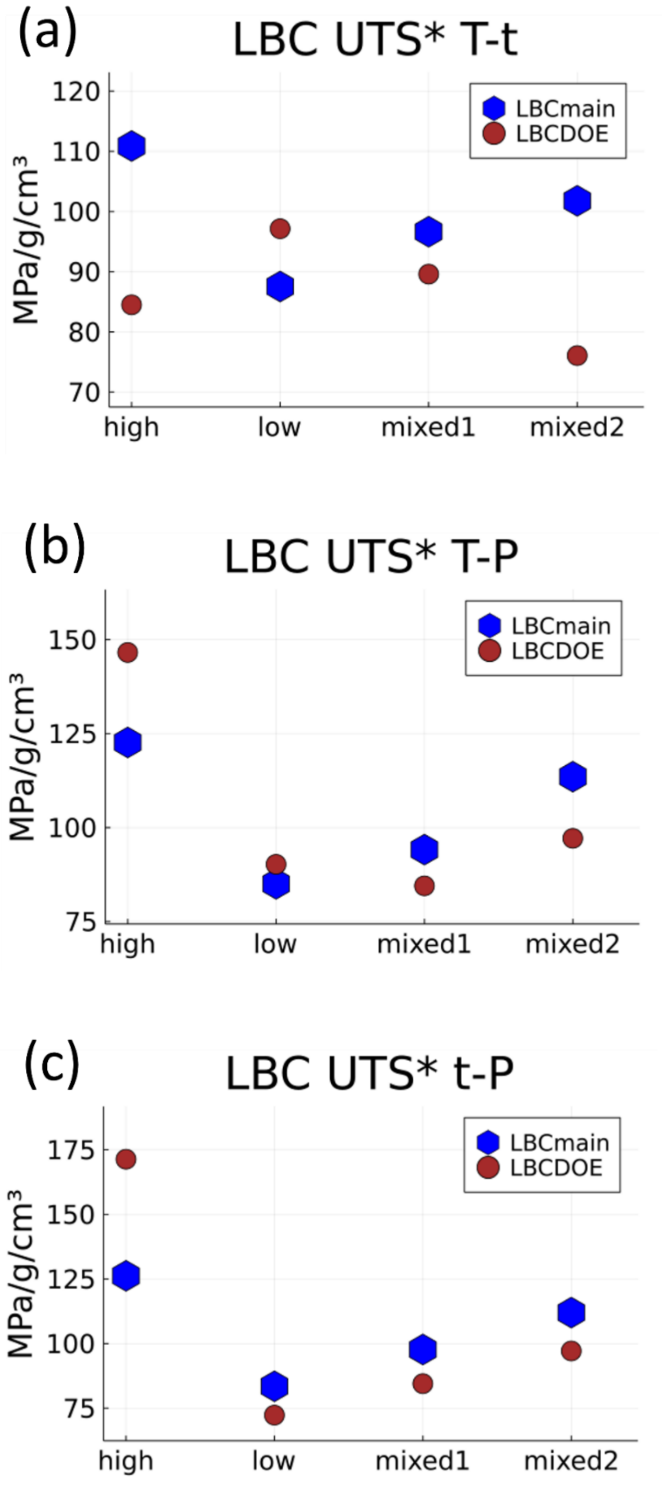


Figure 4.5: LBC UTS* DOE vs. main effects comparisons at different factor-level combinations. Where LBCmain: theoretical combination of main effects, LBCDOE: realized combination of two factors, "high": the property-maximizing combination, "low": the property-

minimizing combination, “mixed1”/“mixed2”: A combination consisting of one factor-level that maximizes and another that minimizes properties, T: temperature, P: pressure, and t: time

4.2.3 Structure-processing-property relationships between BC and LBC

Our DOE analysis revealed that interactions between our processing parameters (pressure, time, temperature) influence the mechanical properties of the LBC composites; in particular, temperature-time and temperature-pressure have synergistic effects. The combination of temperature and time may affect the degree or kinetics of lignin curing and crosslinking or adhesion to cellulose fibrils,[24] while temperature and pressure may be involved with the degree to which deeper-penetrating lignin particles are cured. SEM analysis was conducted on BC and LBC samples to investigate further microstructural differences in response to the processing conditions (**Fig. 4.6**). A common observation across all samples was the formation of defects from the rolling process, evidenced by parallel sheets folding back onto themselves, seen at low magnification (**Fig. 4.6a**). Under low temperature and intermediate pressure and time (120/10/20), we observed discrete lignin particles present in deeper layers of an LBC sample (**Fig. 4.6b**), demonstrating that not all lignin was cured evenly throughout the sample during processing. In agreement with this observation, the lowest temperature, pressure, and time conditions (120/5/10 and 120/10/20) produced the LBC composites with lowest UTS*, EtB, and Tgh*.

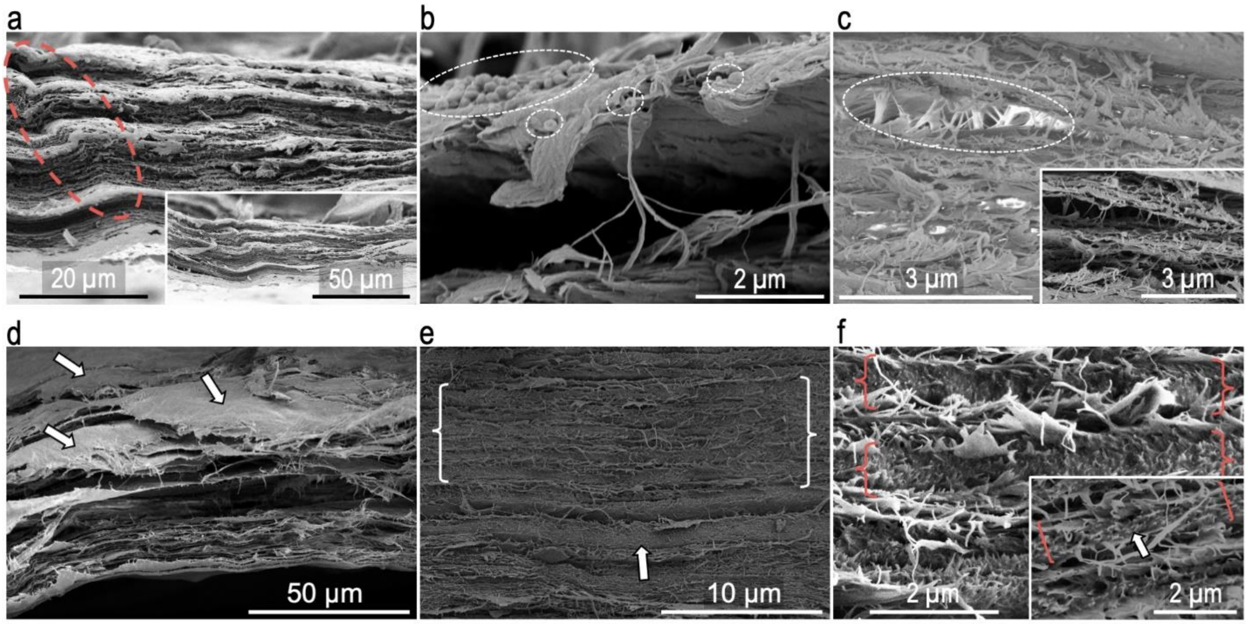


Figure 4.6: SEM analysis of BC and LBC samples. (a) Low-magnification (2000x), cross-sectional view of BC (160/10/10). A defect is highlighted in which the sheet rolls back onto itself due to rolling process before controlled pressing. Inset: A 1000x view of the defect and BC sheet. (b) Uncured, round lignin particles on top of BC fibrils and layers (from 120/10/20). (c) Interlayer binding fibrils highlighted. Inset: The same image in another detector mode to highlight planes and the spacings between them. (LBC sample from 120/5/10) (d) Top and cross-sectional view of a BC sheet showing both amorphized sheets (arrows) and sheets made of discrete fibrils. (BC sample from 120/5/10) (e) An LBC sample (160/5/30; 5000x magnification) with stacked lignin-BC composite planes in brackets. The arrow points at a single composite plane instance. (f) A magnified image of (e; 20000x) highlighting the composite planes and the spacing between each plane. The inset depicts a 50000x magnification to highlight fused and fractured lignin-BC fibrils.

The pure BC material, before any processing, is formed by the sequential deposition of cellulose fibrils in a thick multilayered arrangement as shown in **Fig. 4.1c**. Each layer (hereafter denoted “plane”) is essentially a continuous, porous matrix with discrete cellulose fibrils, which serve as the load-bearing elements. In all the samples (BC and LBC) we observed interplanar fibril bridging (**Fig. 4.6c**), whereby the fibrils from one plane extended into a neighboring one, acting as a primary strengthening mechanism in agreement to prior literature observations.[14,134,135]

As a result of our processing conditions, in addition to the above, amorphized sheets consisting of fused, indistinguishable cellulose fibrils were formed (**Fig. 4.6d**). We speculate that the fibrils fused together during processing due to cellulose layer interactions under elevated temperature-time-pressure conditions, and that the presence of extensive amorphized sheets would hinder individual fibril sliding as a deformation mechanism. From the mechanical property results (**Fig. 4.2**), we saw that BC samples pressed at 160 °C have the lowest average EtB and Tgh* compared to any of the other pressing conditions, regardless of the time. At 30 minutes of pressing time, the samples pressed at 120 and 140 °C have the second lowest EtB and Tgh*. SEM images show the formation of substantially more amorphized layers in the samples with minimized toughness and EtB (**Fig. 4.6d**). On the other hand, the lowest temperatures (120 and 140 °C) at the lower pressing times had the highest Tgh* and EtB. These results indicate that (i) temperature was the primary contributor to amorphization but required an adequate amount of time to optimize the extent of amorphization, and (ii) extensive amorphized layers restricted chain mobility, thus causing reductions in toughness and EtB.

In the composite papers, SEM images revealed an additional plane element which consisted of the base porous cellulose matrix plane densified by the penetration and curing of lignin to form a lignin-BC composite (**Fig. 4.6e-f**). These composite planes showed no distinct interfaces between the lignin filler and cellulose matrix due to the space-filling properties of the solubilized lignin before curing. As in the case of the amorphized sheets, fibril sliding would be impeded in the presence of the lignin binder, at least until failure of the binder. Indeed, we observe a higher average E and E* of LBC compared to BC (**Fig. 4.1a**), supporting that in the elastic regime, before any plastic deformation of lignin or cellulose, the presence of lignin improves binding between adjacent cellulose fibrils to more evenly distribute load.

The best performing samples for pure BC depended on the property of interest, with 140/15/10 having the greatest E* and 140/5/20 and 140/15/10 being the top performers in UTS*, EtB, and Tgh*. The intermediate temperature causing the best properties suggests that there is a temperature threshold differentiating the effects of amorphization. At that threshold (140 °C), the formation of localized amorphized areas begins, and the samples see positive effects in their stiffness, strength, and toughness, possibly due to enhanced intralayer interactions facilitated by the small-sized lignin particle formation which allows load transmission between neighboring cellulose chains, while the interplanar fibril bonding is not yet hindered. As processing temperature and time increase (140/10/30, and all 160°C conditions), the amorphized areas expand and hinder long-range fibril interactions, effectively reducing strength, stiffness, and toughness, which is what our mechanical testing results show. In addition, at those conditions thermal degradation may start to occur, contributing to mechanical property deterioration.

Regarding the LBC composites, the best performance in any of the mechanical properties consistently came from conditions 120/15/30, 140/15/10, and 160/15/20. As concluded from the DOE analysis presented in the previous section, the performance of the LBC composites is highly dependent on the interactions between the processing factors, rather than on a single factor. We note, however, that the composites with minimum strength and toughness are those formed at the lowest temperature, time, and pressure levels (120/5/10 and 120/10/20), which also contained uncured lignin particles (**Fig. 4.6b**). When comparing the performance of LBC composites to that of pure BC processed at 120 °C, we see that lignin caused a substantial increase in the E, with no effect on UTS and an overall reduction in EtB and Tgh. Yet, at the longest pressing time at that temperature (120/15/30), the stiffness, strength, and toughness are all improved in the presence of lignin. Therefore, lignin curing can improve the overall mechanical properties of composites below the amorphization threshold for the pure BC.

4.2.4 Wetting behavior of BC and LBC composites

Natural lignin is known to promote water hydrophobicity in wood and therefore minimize swelling that could structurally degrade wood. However, the extraction and processing of lignin from woody biomass significantly impact the water properties (solubility and hydrophobicity) of lignin products through changes in the molecular weight and the degree of sulfonation.[136] We conducted water droplet tests to measure the effects of processed, sulfonated (water-soluble) lignin in our composites. By utilizing a combination of a 3-parameter decay function and the Avrami model proposed by Farris et al.,[137] water absorption properties were further analyzed. Contact angle measurements were taken over a period of 2 minutes, allowing for the differentiation between water absorption versus spreading. These two phenomena are assumed to occur simultaneously.

The 3-parameter decay function used here is:

$$\theta(t) = \theta(0)e^{kt^n},$$

where θ is the contact angle measure in degrees, t is time in seconds, and the parameters k and n represent the contact angle evolution rate and the relative fraction of absorption versus spreading (0 = pure absorption, 1 = pure spreading), respectively. Compared to the strongest BC sample (140/5/20), the weakest BC (160/5/30) had a higher immediate contact angle (83.7° versus 76.5°) but showed more absorption of water at a significantly faster rate (**Fig. 4.7a-b**). The strongest LBC composite (120/15/30) had a lower immediate contact angle compared to the weakest (120/5/10) (40.7° versus 57.8°), but showed roughly equal absorption and spreading between the two samples, demonstrating that processing conditions do not affect the long-term water resistance in the composites. The results show that the incorporation of lignin imparts hydrophilicity on the material, as evidenced by the lower contact angle values, which is justified from the use of water-soluble lignin in these composites.[136] As discussed earlier, SEM images revealed uncured lignin particles in the composite with weakest mechanical properties (**Fig. 4.6b**), while the strongest composites had a continuous lignin-infused matrix. Given the hydrophilicity of our lignin, the lower contact angle for the composite with more even lignin spreading (strongest sample) is justified.

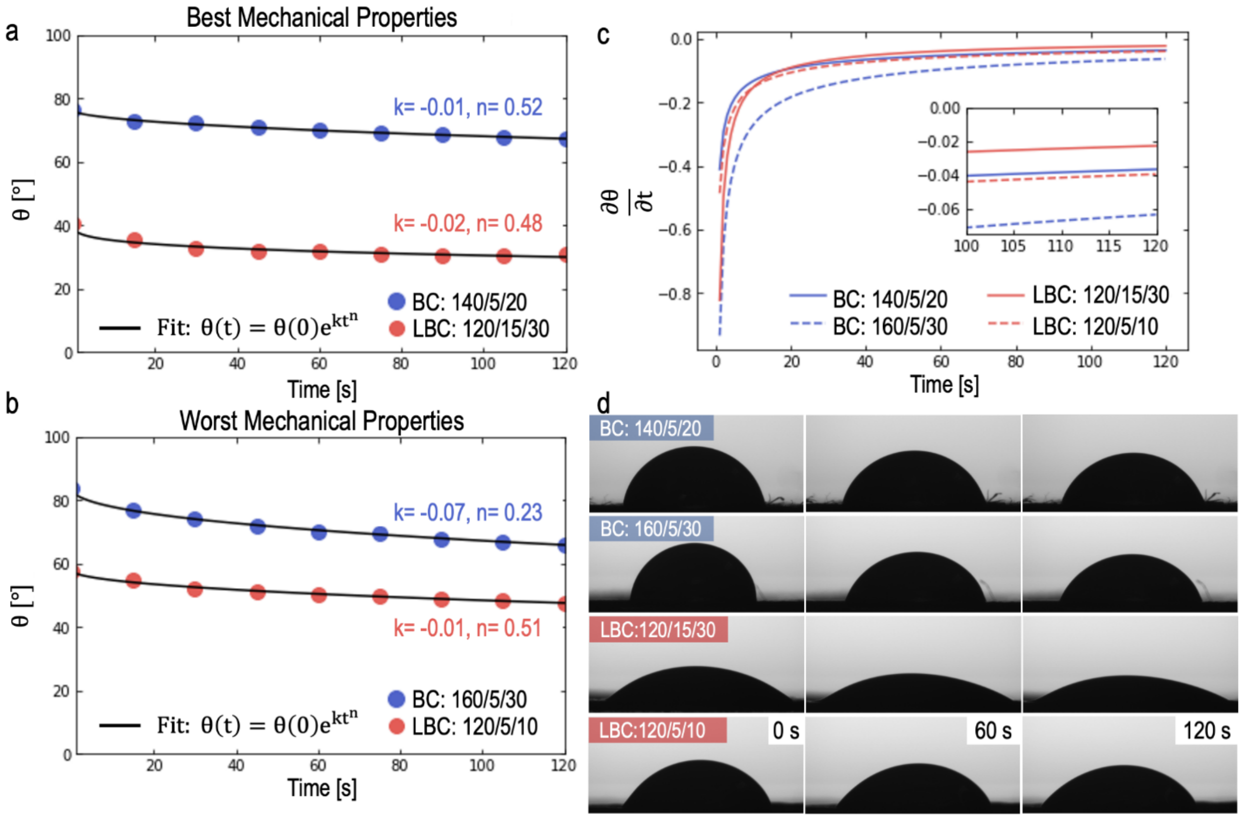


Figure 4.7: Water absorption in BC (blue) and LBC (red). (a) Time-dependent contact angle and fitted curve for the best mechanically performing BC and LBC sheets. (b) Time-dependent contact angle and fitted curve for the worst mechanically performing BC and LBC sheets. (c) First derivative of the fitted data with inset magnifying the last 20 seconds. (d) Images of water droplets at 60 s time intervals.

Of the four studied samples, the strongest and weakest LBC have a similar average surface roughness (P_a) of 4.4 ± 2.4 μm and 4.4 ± 2.0 μm, respectively. In contrast, the strongest and weakest BC have a P_a of 3.7 ± 0.4 μm and 6.8 ± 3.2 μm, respectively. These results confirm the hypothesis that the highest surface roughness (weakest BC, k = -0.07, n = 0.23) displays the greatest water absorption over time and highest initial hydrophobicity (at t = 0 s). By the end of two minutes, there is still absorption and spreading of water occurring, most prominently in the weakest BC sample. This observation is emphasized by the first derivative of the contact angle over time (generated from the fitted data, **Fig. 4.7c**). While none of the samples reached the zero-speed rate within the two-minute time frame, the weakest BC sample starts and continues to wet the fastest. Snapshots of the water droplets at 0 s, 60 s, and 120 s are compared in **Fig. 4.7d**.

4.3 Conclusions

In this chapter, we presented the effects of incorporating lignin in BC matrices and analyzed how altering processing (cold- and hot-pressing) temperature, pressure, and time conditions tunes the mechanical properties in the resultant composite papers. Infiltrating lignin in the layered cellulose matrix comes with property trade-offs: a significant improvement in stiffness at the cost of creation of new sources of defects for brittle failure. Our results demonstrated that there is a temperature-pressure and temperature-time threshold above which lignin penetration into and incorporation in the dense BC structure is optimized to form a continuous composite matrix with enhanced stiffness, strength, and toughness compared to the base BC matrix. Below that threshold, lignin particles are not able to infuse into the BC matrix, causing an enhancement to only the Young's modulus while reducing improvements to strength and toughness.

We reported a cellulose amorphization mechanism controlling the mechanical behavior of pure BC materials in response to the processing conditions. When the processing temperature and time are intermediate (140 °C) the formation of amorphized intralayer areas, in which cellulose fibrils fuse together, enables an improved strength, modulus, and toughness. However, at higher temperatures and processing times, the possible cellulose degradation as well as the enlargement of amorphized layers, which hinders discrete intra- and inter-layer fibrillar interactions, effectively reduce the strength and toughness of pure BC papers.

Finally, we investigated the effects of processing on our nanopaper hydrophobicity, given that lignin confers hydrophobic properties in nature. We found that the same solubility that allows lignin to infiltrate the BC matrix and improve mechanical properties causes the BC matrix to become more hydrophilic. That is, the lignin in the best mechanically performing sample was better integrated throughout the matrix, causing greater absorption and spreading of water initially and over time. The opposite behavior was found for the worst mechanically performing sample, implying that lignin was less well incorporated and distributed in the BC. Our results demonstrated how introducing lignin and varying processing conditions allow different mechanical and water properties to be achieved in BC papers.

5. Using machine learning to guide the synthesis of multiscale, multicomponent bionanocomposites

5.1 Context for the third chapter based on lessons learned from the first two chapters

From Chapter 1 (3D-printed spirulina nanocomposites), we found that drying methods controlled microscale morphologies and resultant stability of the macroscale spirulina composite. In particular, oven-drying (OD) and desiccator-drying (DD) resulted in cracked samples while freeze-drying (FD) resulted in uncracked samples. OD and DD also resulted in the densest and smallest volume samples, supporting that the macroscale cracking was due to nonuniform warping from the drying process. This warping could be mitigated by the addition of micro-crystalline cellulose fibrils (CF); adding as little as 1 wt.% CFs into the spirulina slurry stopped cracking in the final composites while allowing uniform shrinking.

SEM micrographs revealed the processing-dependent effects on microscale morphology. For example, DD at ambient temperature caused densification and amorphization of the spirulina biomass, leaving large voids in the microstructure as material was pulled together unevenly. The amorphous matrix consisted of protoplasmic cell contents and cell wall fragments as the drying caused cells to rupture. In contrast, OD introduced heat-induced gelatinization (matrix plasticization with water) as a structure-forming mechanism. As a result of whole, unruptured cells fusing to the amorphous matrix from the gelatinization, a bulk blend morphology that resulted in a toughened composite was formed. Finally, FD resulted in porous foam structures due to the sublimation of ice crystals uniformly distributed throughout the biomass slurry. This sublimation occurred without rearrangement of cells and cell components, allowing for the formation of micropores between intact cells.

We found that crazing was a toughening mechanism present in samples with an amorphous matrix (DD, OD), with and without the bolstering of cellulose polymers. CFs were well-integrated into the matrix as well, forming flake and cluster structures that were fused with the surrounding matrix. We showed that the spirulina system is compatible with other biofillers despite not containing any cellulose itself, making it amenable to combination with cellulose and potentially other hydrophilic biopolymers.

In the second chapter (processing of lignin-BC composites), we used a design of experiments (DOE) approach to correlate processing conditions to mechanical and water-repellency properties in lignin-bacterial cellulose (BC) composites. The Latin squares design minimized the number of combinations of processing conditions we had to test, but also confounded the conditions so that we had to infer interactions between each to draw conclusions. From our analyses, we found that the processing conditions were largely independent in pure BC (a single component system) but interacted when lignin was incorporated into the BC matrix (a two-component system). From SEM analysis, we discovered the processing-induced formation of three separate structures in the composite: 1) amorphized sheets, whereby BC fibrils fuse together and act as a chain-sliding inhibiting unit; 2) uncured lignin-BC composites, whereby lignin particles nucleated or diffused between BC layers but did not fuse with the BC; and 3) cured lignin-BC composite, whereby lignin impregnated the BC matrix and fused so that a continuous lignin-BC entity was formed. We also discovered that it was possible to impose conditions that cured lignin while minimizing or preventing BC amorphization.

Besides the formation of these three new structures, adding lignin to the BC did not fundamentally alter the layered BC microstructure, even after processing the composite under heat, pressure, and pressing time conditions. Literature and our results suggested that BC fibrils act as the primary load-bearing elements under stress, with interfibrillar hydrogen-bonding promoting a toughening mechanism known as chain sliding. The presence of cured lignin-BC structures and/or amorphized BC structures resulted in strengthening and embrittleling of the composite as this chain sliding mechanism was inhibited.

Finally, we found that micromorphological alterations had implications in the water-repellency properties of the composites as well. The weakest BC composites (ones that did not contain amorphized sheets) were found to have a greater surface roughness and promoted greater absorption of water than the strongest BC composites that contained fewer pores due to the presence of amorphized sheet elements. In contrast, the weakest lignin-BC composites (ones that were processed under conditions that did not cure the lignin) resulted in greater hydrophobicity than the strongest composites due to the non-uniform distribution of the hydrophilic lignin. The strongest lignin-BC composites, whose lignin was well-incorporated in the BC matrix, was found to be more hydrophilic as the lignin absorbed more water throughout the entirety of the base matrix.

In the third and final chapter, we created a machine learning model to predict the properties of a composite consisting of similar materials to the ones used in the previous two chapters. The goal of the machine learning model was to maximize a target mechanical property given the compositional landscape of a 3-component composite consisting of algal biomass, bacterial cellulose, and a biobased molecule. The homogenizing and processing conditions of the composites were kept uniform while only differences due to dry composition were investigated. By confining the investigation to only composition, the upper and lower bounds of the input variables were known, ensuring that a global optimum could be found for the given processing conditions. Further investigations would apply the optimal composition to varying processing conditions.

5.2 Materials and Methods

See **Chapter 2** for comprehensive details on the methods used. In brief, we used SEM to image the surface layers and fracture cross-sections of composites and rotational viscometry to measure the viscosity of suspended composite slurries after blending and before casting. Tensile testing was performed to understand the mechanical properties (Young's modulus, strength, elongation to break, and toughness) of composite film strips.

The biomatter material used for this chapter's work is *Chlorella vulgaris*, a unicellular spherical microalgal species. We used in-house cultured bacterial cellulose (as described in **Section 2.2.3 Methods**) as the fiber element of the composite. Stearic acid was used as the third component in the ternary mixture.

5.3 Introduction - Machine learning definitions and fundamentals

5.3.1 Algorithms

ML algorithms can be broadly categorized into several classes depending on the purpose and characteristics of the data.[67,138] If the relationships between data are unknown, *unsupervised learning* may help identify patterns and structure in the data. However, in materials science, inputs often have well-defined outputs (e.g. processing leads to properties) and thus the data is "labeled". Labeled data can be processed using *supervised learning* algorithms. Under supervised learning, *classification* methods can be used to distinguish categorical data (e.g. is this material a metal or polymer?) while *regression* is used to predict output quantities based on similar inputs (e.g. similar processing conditions have the same underlying physics and so should yield similar properties). This algorithm categorization is summarized in **Fig. 4.1**.[138]

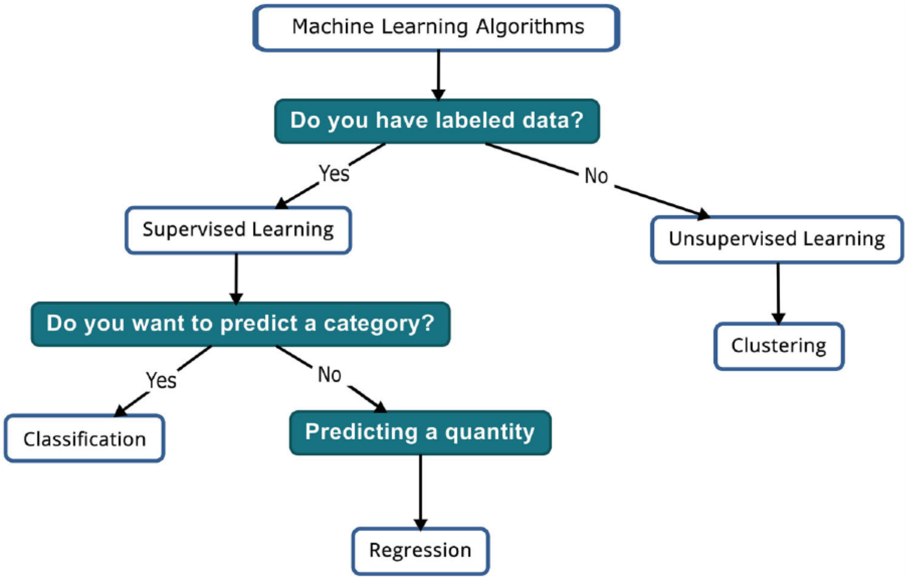


Figure 5.1: Machine learning algorithm selection summary. Materials science predominantly uses regression or unsupervised learning methods. Adapted from Gomez et al.[138]

Regression algorithms are of primary interest because they can be used to predict properties from structure or processing information.[67,139] The relationship between the input information and the output properties is described by the *model* that is produced by the algorithm.[73]

5.3.2 Machine learning terms and definitions

There are several components within a regression algorithm that are required to build a model. *Features* are inputs in the model that describe the data in some way and feature engineering describes a process by which informative features are selected. Informative features are features that have high predictive power of the model output. Given inputs and outputs, a model is made that minimizes the *loss function*, which is a function that quantifies the difference between actual data points and predicted data points in the model – in statistical regression terms, the loss function describes the data error. As more inputs are given to a model (either through the addition of more data or features) the model updates according to a defined *optimization* function or routine that tells the model how to update. The implementation of these functions is often automated (or can be specified) in the algorithm itself in a ML library or platform, and so are not further discussed here.

The *degrees of freedom* in a ML model refers to the complexity of the model. Complexity increases with the number of fitting parameters. A model with more fitting parameters may describe data better than one without as many parameters and is said to have a greater total degrees of freedom. However, more complex models usually come at the cost of overfitting, which means that the model fits the given data (the training set) well but does not generalize to fit new values well. In ML vernacular, the model has low bias and high variance. There is a well-known bias-variance tradeoff in ML, where models that generalize well (low variance) do not fit the data particularly well (high bias). Thus, the precision of a model is dictated by this tradeoff. Here, we define *precision* as how closely the model predicts future data values, including uncertainty.

A recent paper demonstrated that there are two ways to increase model precision: adding more data and increasing degrees of freedom.[71] Adding data obtained from less precise or accurate methods, or from similar but not the same experiments as currently being conducted, improved the precision of the model while minimizing the costs of obtaining new data. For example, the data could come from simulations or physical experiments using related experiments. Not surprisingly, increasing degrees of freedom increases the complexity of the model and leads to a better fit to the training data, but may not generalize well when trying to predict new data values. To avoid adding unnecessary complexity to a model, *regularization* terms are often added to a model's loss function. Regularization terms are additional hyperparameters in a model that penalize increasing complexity (i.e. greater coefficients on model terms), encouraging the model to fit a simpler curve. Being able to add to a dataset from related sources opens avenues to increase model precision while retaining generalizability from not having to introduce additional, potentially unnecessary, features or parameters to the model.

Some algorithms use non-parametric methods to classify and regress data. Non-parametric methods try to guess a specific function that fits the data rather than parameters/coefficients for an assumed or known function. Rather than relying on an assumption about the relationship (e.g. linearity) between datapoints a priori, non-parametric methods use *kernel functions* that define a notion of "closeness" or "similarity" for input data. A kernel function describes characteristics of the functions that will describe/fit the data. There are different kernel functions each with their own characteristics, but kernels may also be added and multiplied together to create new kernels with aspects of their parent kernels. Two common kernel functions are the periodic and radial basis function kernels. As the name implies, the periodic kernel ensures that the suggested functions will be periodic in nature. The radial basis function (RBF) kernel weights data points in such a way that neighboring data points influence the values of each other. So, data points that are right next to each other will have similar values, ensuring that potential fitting functions will look smooth instead of jagged (e.g. linear in between each given data point).

Any two points in a fitting function described by a kernel are related to each other through a *covariance matrix*. Each term in a covariance matrix is defined by the kernel (AKA covariance) function evaluated at each pair of points. The covariance function for an RBF kernel is the following:

$$K(x, x') = \sigma^2 exp\left(-\frac{\|x-x'\|^2}{2l^2}\right) \tag{3.1}$$

where x and x' are any two points in the fitting function. To distinguish from the input data in a model, the term *hyperparameter* is used by convention for the other alterable parameters in a covariance function, and also in ML models in general. In the RBF kernel, σ^2 is a hyperparameter for the variance that influences how far the function can deviate away from the specified function mean (thus determining the vertical scale of the output). The hyperparameter *l* determines the horizontal length scale over which points are correlated. More strongly correlated points are more strongly influenced by each other. For example, if we have 100 data points and a large *l*, each data point is strongly influenced by each other, meaning that their output values, f(x), will all be close to each other, and the output function will look very smooth. In the opposite case where *l* is small, only the nearest neighboring data points influence each other, resulting in potentially rougher distributions. For example, in the first case, a high $f(x_{40})$ would imply a high $f(x_{10})$, while in the second case, the $f(x_{40})$ value would have little influence on the value of $f(x_{10})$ and vice versa.

As an aside, the covariance function can be generalized to multiple dimensions, or input variables in the scientific experiment sense. The covariance function describing the relationship between a variable of interest and all other (input) variables would be, for the case of an RBF kernel:

$$K(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 exp\left(-\sum_{j=1}^{N} \frac{\left\| x_j - x_j' \right\|^2}{2l_j^2}\right) \tag{3.2}$$

where $\boldsymbol{x}$ and $\boldsymbol{x}'$ are matrices of vectors x_j of input values for each dimension *j*, and *N* is the number of dimensions/variables. The general form of the covariance function does not change except to include contributions from each input variable.

Before using a model on new data, old data must be used to train and condition the model. This old data is known as the training set. However, the selection of the training set itself varies the hyperparameters of the model, necessitating the *validation* of the model before use on real data. For this process, a type of bootstrapping protocol called *cross-validation* is commonly used. In cross-validation, a model is fit with a large subset of the data and then the model fit is tested on the remaining data. The model is then fit with another subset of the same data and tested on the new remaining data. The model is iteratively fit until an average fit of the parameters can be found.

5.3.3 Support Vector Machines and Gaussian Process Regression

Two regression algorithms that have gained adoption in materials science are Support Vector Machines/Regression (SVR) and Gaussian Process Regression (GP). Both algorithms produce a model that describes a non-parametric function fit to the data. That is, there is no predefined form of the function (e.g. linear, exponential, polynomial, etc.) and that function is updated as more training data is fed to the algorithm via Bayesian inference.

SVRs output a curve fit defined by data that fall within set boundaries. The hyperparameters for the boundaries define how tight-fitting the boundaries are around the curve and how much the model is penalized for having data that fall outside of the boundaries.[140] **Fig. 5.2a** illustrates this concept with an example and explanation of the hyperparameters.

GPs fit the data to an infinite family of functions, using the training data as pinning points that the functions must pass through. New data is incorporated into the training set as they are collected, which increases the number of pinning points and reduces the number of possible functions that fit all of the data. GPs assume each data point is a realization of a normal distribution with mean μ and (co)variance $\Sigma = K + \sigma^2$, where σ^2 is the usual variance about mean μ and K is the covariance/kernel function. μ is the data point value (or its expectation value for unrealized points) that the GP fitting function passes through and σ describes the uncertainty about each data point. Uncertainty permits a fitting function to take on a value close to, rather than exactly at, a realized data point. Finally, K describes the curve that fits the rest of the function.

Potential fitting functions about unrealized points can be seen as random samples drawn from the normal distribution about the true data mean μ, meaning a confidence interval is naturally formed about every point as a consequence of the assumptions of GPs. Once a value is found, the confidence interval narrows down at that point because the exact mean (within error) of the normal distribution at that point is determined. This confidence interval can be used as a mathematically informed method to selecting future experiments, by conducting experiments at parameters where uncertainty is highest and most information can be gained. **Fig. 5.2b** illustrates this concept.

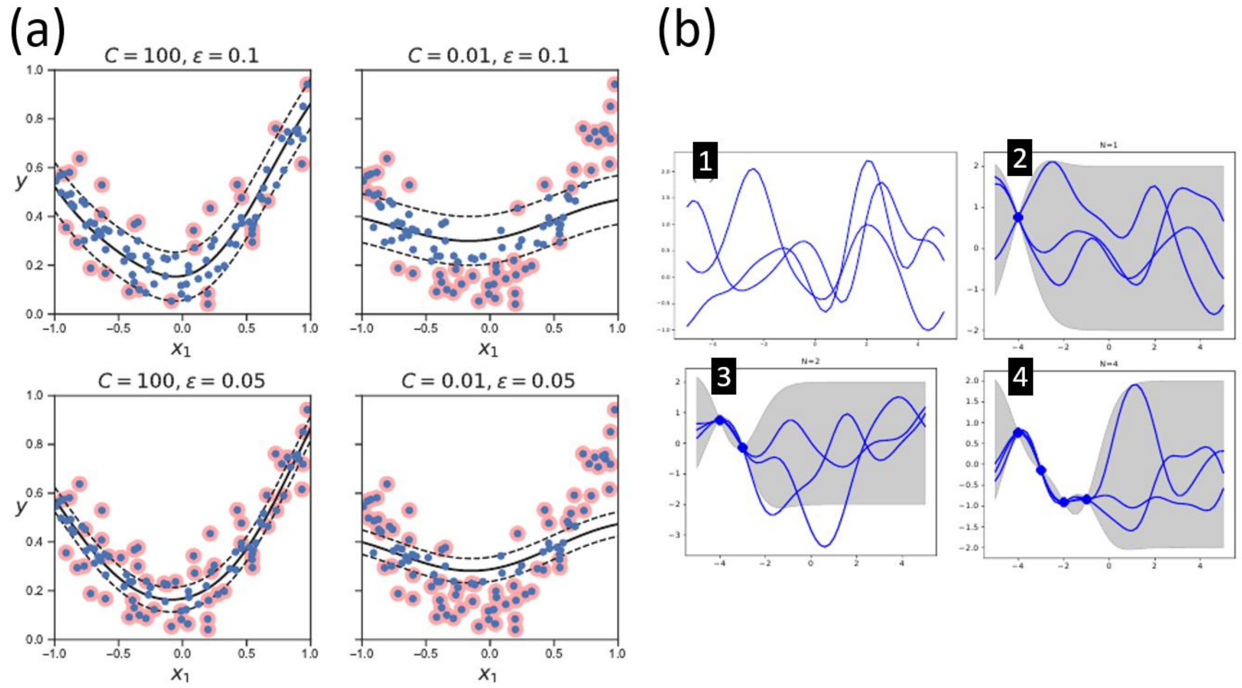


Figure 5.2: Example illustrations of SVR and GP regression algorithms
(a) SVR with epsilon insensitive loss function regularization: given a set of data points, parameters *C* and ε control how much the data is manipulated. ε dictates the size of the tube surrounding a fitting function – data points inside of the tube are not penalized. *C* dictates the weight of the penalty for data points that lie outside of the tube. In particular, narrow tubes and high penalty lower variance of the training data but may lead to overfitting.
(b) Illustration of GP model using an RBF kernel based on hypothetical fitting functions. Axes are arbitrary. [1] Representative sampling of an infinite family of potential fitting functions. [2] Collection/Realization of a single data point. The family of functions are narrowed down to those

that pass through the realized data point. The gray area represents the confidence interval of the potential functions. With no additional data points, the confidence intervals are essentially unbounded. [3, 4] Addition of data points. The hypotheses (functions) are further updated to interpolate the realized data points as the function space becomes better defined with the collection of new data. Due to the smoothness assumption from the kernel, the uncertainty between two close points (i.e. data from similar input parameters) may be smaller than two points that are farther apart (i.e. drastically different input parameters).
Both sets of images adapted from Murphy.[140]

SVRs use a frequentist approach to determining the fitting curve called maximum likelihood estimation (MLE). The viewpoint of MLE is to select a curve such that the probability of the output data given the curve is maximized. This maximization occurs by changing the fit parameters of the assumed underlying distribution/curve. In essence, a single curve is produced that maximizes the probability that the output data comes from that curve's distribution (within variability). In contrast, GPs use a Bayesian approach to determining the curve (model) fit called maximum a posteriori (MAP) estimate. MAP produces a hypothesis function that best fits the current data and is likely to fit future ones as well, given prior hypothesis functions that previously fit the data best (or that we guessed based on prior experience). As new data is obtained, rather than changing a parameter on a single assumed function, the algorithm weighs the likelihood of other models fitting current and future data and updates accordingly. The difference between MAP and MLE is that the former uses knowledge about the prior best fitting functions to inform the new hypothesis, while the latter uses all of the data concurrently.

In the next section, we use a GP model over an SVR for the proposed experiments due to the GP's intrinsic inclusion of the uncertainty at unsampled points. Both the data means and the uncertainty over the entire sample space can be used to decide the next experiments using an "exploration-exploitation" strategy.[70] Using this strategy, an algorithm is able to decide whether to (1) [exploit] suggest a next experiment (or experimental conditions) that is similar to an already tried experiment in an attempt to find the local extremum in properties in that area of the experiment space, or (2) [explore] suggest an experiment in the vicinity of greater uncertainty (i.e. vastly different experimental conditions than previously tried) to lower the uncertainty in that portion of the experiment space, in an attempt to find the global extremum over the entire experiment space. GP models scale as n^3 datapoints,[86] making them inefficient for large datasets, but optimal for small datasets such as will be used here. Other supervised algorithms, such as neural networks, may also be considered for use at a later time, but are deferred for now due to their black box functionality of not being able to intuitively explain which features are most important for describing the model/physical phenomena.

5.4 Machine learning model workflow

The framework for the experimental procedures is Bayesian optimization. This framework consists of two main components: a surrogate function, which models the data, and an acquisition function algorithm, which informs the next experimental conditions to probe. The surrogate function will be derived using a GP model as described earlier. The acquisition function, sometimes called a selector,[70] is a separate function that takes the input of the GP model (the surrogate function and its confidence interval) and informs where to gather data next. This function depends on the acquisition algorithm used, which has criteria to determine the "exploration-exploitation" strategy (explained earlier) for the sample space.

Studies have shown that the performance of an acquisition algorithm depends on the nature of the dataset.[70,141] The explanation of the algorithms that exist are deferred to a later discussion, but algorithms that have been used in the materials science field in conjunction with GP models include Expected Improvement (AKA Efficient Global Optimization), Kriging Believer, Knowledge Gradient, Upper/Lower Confidence Bounds, Greedy/Min, and Thompson Sampling.[68,70,80,83,141]

The following workflow was used for setting up and applying a model:

1. Identification of a mechanical property as a target parameter for the 3-component composite: absolute toughness (Tgh) and elongation to break (EtB).
2. Refinement of features to only dry composition.
3. Collection, input, and updating of data into a regression model (gaussian process) with a simple RBF kernel with variable length scale.
4. (Acquisition) Algorithmic selection of next parameters/experiments to attempt to optimize target property or combination.
5. Collect data with suggested experimental parameters.
6. Repeat steps 3-5 for as long as required.

5.5 Results and Discussions

5.5.1 Experimental set-up

To confine the compositional space in the model, pinning point compositions were selected at extrema and pseudorandom points inside of the compositional space. **Fig. 5.3** depicts the selected compositional pinning points.

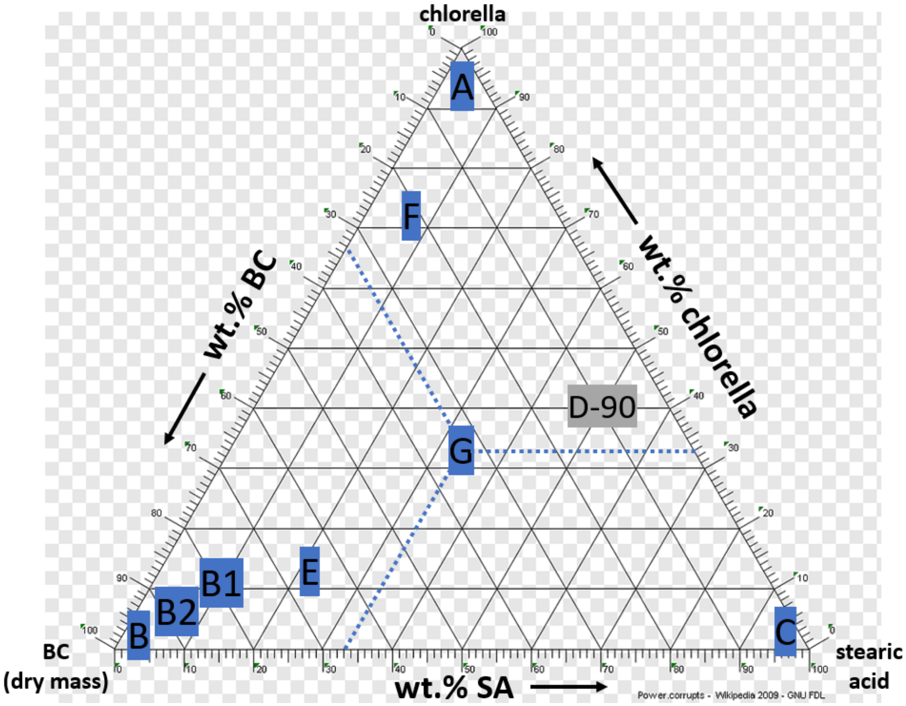


Figure 5.3: Ternary compositional map of chlorella-(dry) BC-stearic acid composites. Dashed lines drawn from experiment/pinning point G are representative of the appropriate axes to read from any of the points to establish the composition of the respective experiment.

Pinning points A, B, and C were selected such that the highest concentration dry component was 99 wt.% of the composite. While successfully cast and dried, these experiments showed varying levels of robustness. Experiments A and C resulted in films that were too brittle to test upon removal from the casting tray; experiment B remained intact. **Table 5.1** summarizes the compositions of each of the experiments.

Table 5.1: Compositions of training and test data points in chlorella-BC-stearic acid composites

Name of experiment/pinning point	Chlorella wt. %	BC wt.%	SA wt.%
A	99	0.5	0.5
B	0.5	99	0.5
B1	10	80	10
B2	5	90	5
C	99	0.5	0.5
D-90	40	10	50
E	15	65	20
F	73	20	7
G	33.3	33.3	33.4

5.5.2 The effect of absolute solvent mass to dry mass ratio

The D series of experiments (40 wt.% chlorella, 10 wt.% BC, 50 wt.% SA) were selected as a set of screening experiments to determine whether absolute liquid content had an effect on the formation of the casted films. Formulations ranging from 45 mg total solvent (D-45) to 450 mg total solvent (D-450) were synthesized. Owing to the density of ethanol being less than 1 g/mL, D-450 exceeded the capacity of the blending cup and had to be partially discarded. D-450, D-360, D-270, and D-180 were able to be cast, but the resultant films were too thin to remove and test, given the physical constraints of the container. Qualitatively, the samples cast and dried differently. **Fig. 5.4** depicts the qualitative differences between dried D-45, D-90, and D-180 experiments. While D-180 was uniform in thickness, it was thin and adhered to the tray, making removal of the film difficult; no samples could be recovered for testing. D-45 showed areas of non-uniformity in addition to warping after drying. The warping could be a result of uneven drying stresses due to non-uniform distribution of material or uneven thicknesses, as elaborated in **Section 1.5.** D-90 dried as a flat sheet, indicating an optimal ratio of liquid mass to dry mass. Further investigations of the effect of absolute liquid volume on drying characteristic were deferred to a future study. All other experiments in this study added the minimum amount of ethanol necessary to retain a 1.6:1 ethanol mass to (BC) water ratio. This ethanol ratio was selected to ensure complete dissolution of free stearic acid, which is otherwise insoluble in water.

D-45 D-90 D-180

Figure 5.4: Pictures of D series experiment samples. The number indicates the absolute amount (by mass) of solvent added to the composite mixture before drying.

Table 5.2 briefly summarizes the results of tensile testing on D-45 and D-90. In addition to warping less, D-90 is approximately 3 times stiffer and 1.5 times stronger than D-45. Density was similar in both samples at 0.4 g/cm^3.

Table 5.2: Mechanical property comparisons of D-45 and D-90

Experiment	**YM (MPa)**	**Strength (MPa)**	**EtB (%)**	**Density (g/cm^3)**	**Sp. YM (MPa/g/cm^3)**	**Sp. Strength (MPa/g/cm^3)**
D-45	111 ± 37	1.6 ± 0.4	2.3 ± 0.8	0.43 ± 0.05	279 ± 122	4.2 ± 1.3
D-90	362 ± 81	2.8 ± 0.6	1.1 ± 0.2	0.39 ± 0.05	912 ± 192	6.8 ± 1.7

5.5.3 Viscometry

The viscosities of the slurries were measured by making excess slurry and setting aside a portion to be measured in a viscometer. **Fig. 5.5** summarizes these results. Each sample showed shear thinning behavior, despite the inability of some samples to achieve full homogenization from the blending protocol. Qualitatively, many of the mixtures had chunks of BC that were visible with the naked eye that conferred a lumpy texture to the slurries. Interestingly, experiments A and C, the experiments whose samples could not be tested due to brittleness, showed the lowest viscosities at all shear rates. Non-extrema compositions tended to show higher viscosities, implying an interaction between the components. There was no discernible relationship between the content of specific components and their absolute viscosity measurements or the rate at which they thinned.

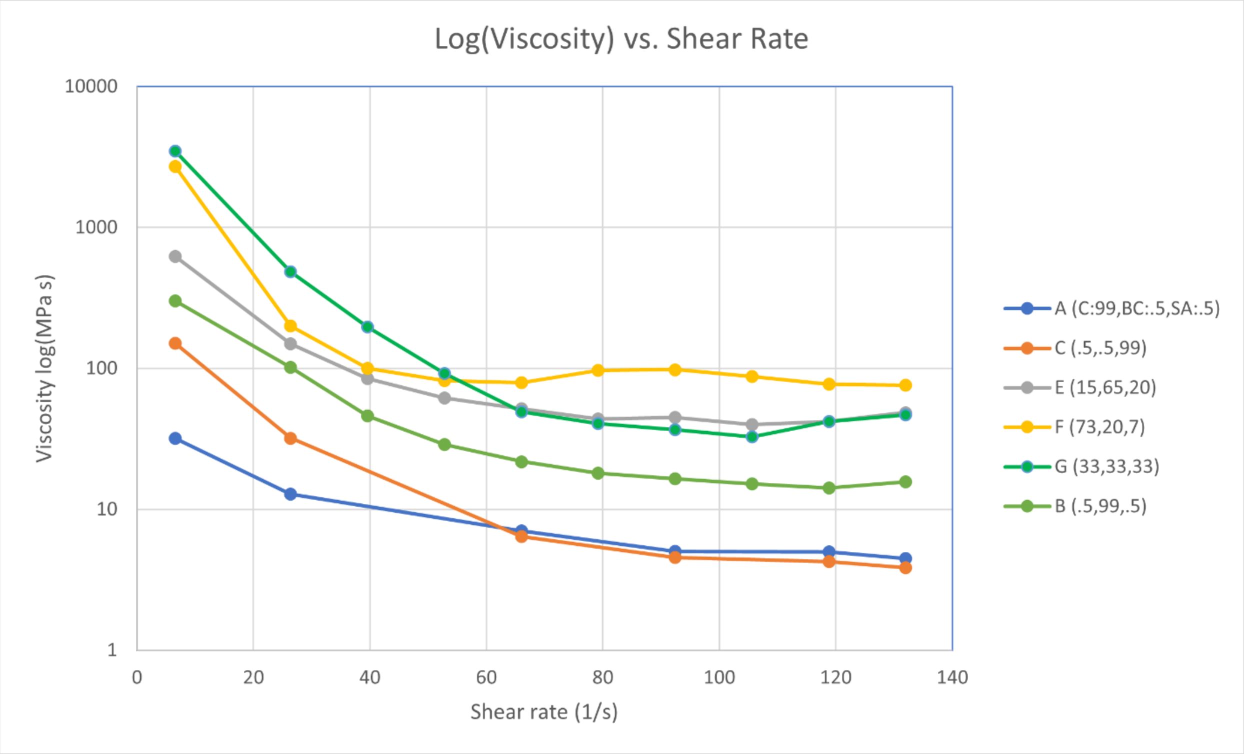


Figure 5.5: Viscometry data. The viscosity of small volumes of homogenized sample was measured at varying shear rates. All samples showed shear thinning behavior to a varying degree.

5.5.4 *Tensile testing*

Other compositions were successfully synthesized with varying degrees of warping. Tensile specimens of ASTM D1708 “dogbone” shape were stamped out and tensile tested for each of the samples. **Fig. 5.6** depicts boxplots of various mechanical properties and **Fig. 5.7** stress-strain curves for each experiment. Both plots reveal the presence of outliers in some of the samples. For the machine learning data set, clear outliers were removed to facilitate accurate predictions for compositions.

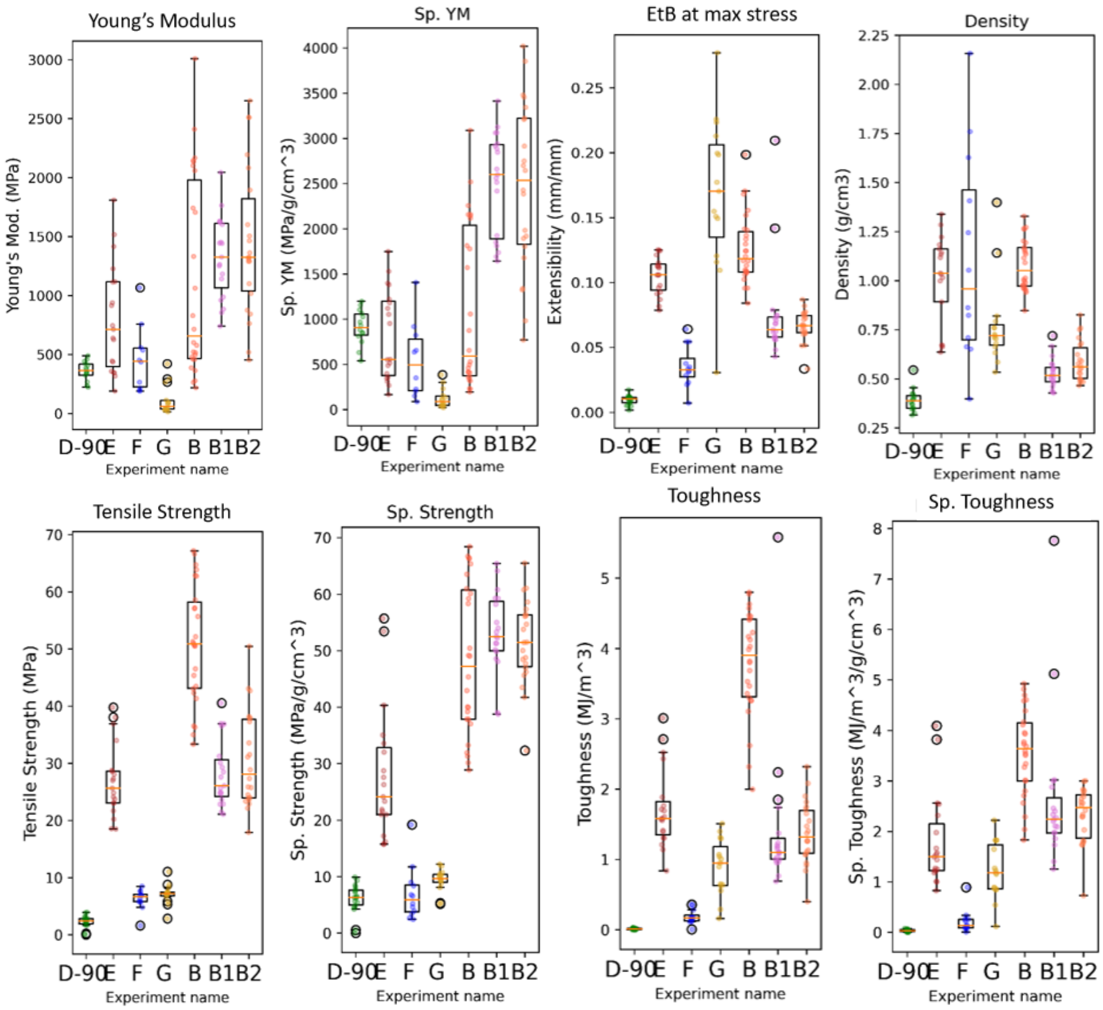


Figure 5.6: Boxplots of mechanical properties of experiments. These boxplots represent all data collected from tensile testing, including outliers.
Key:
Sp. = specific (normalized to each sample’s density measure)

YM = Young’s modulus
EtB = Elongation at max stress before break

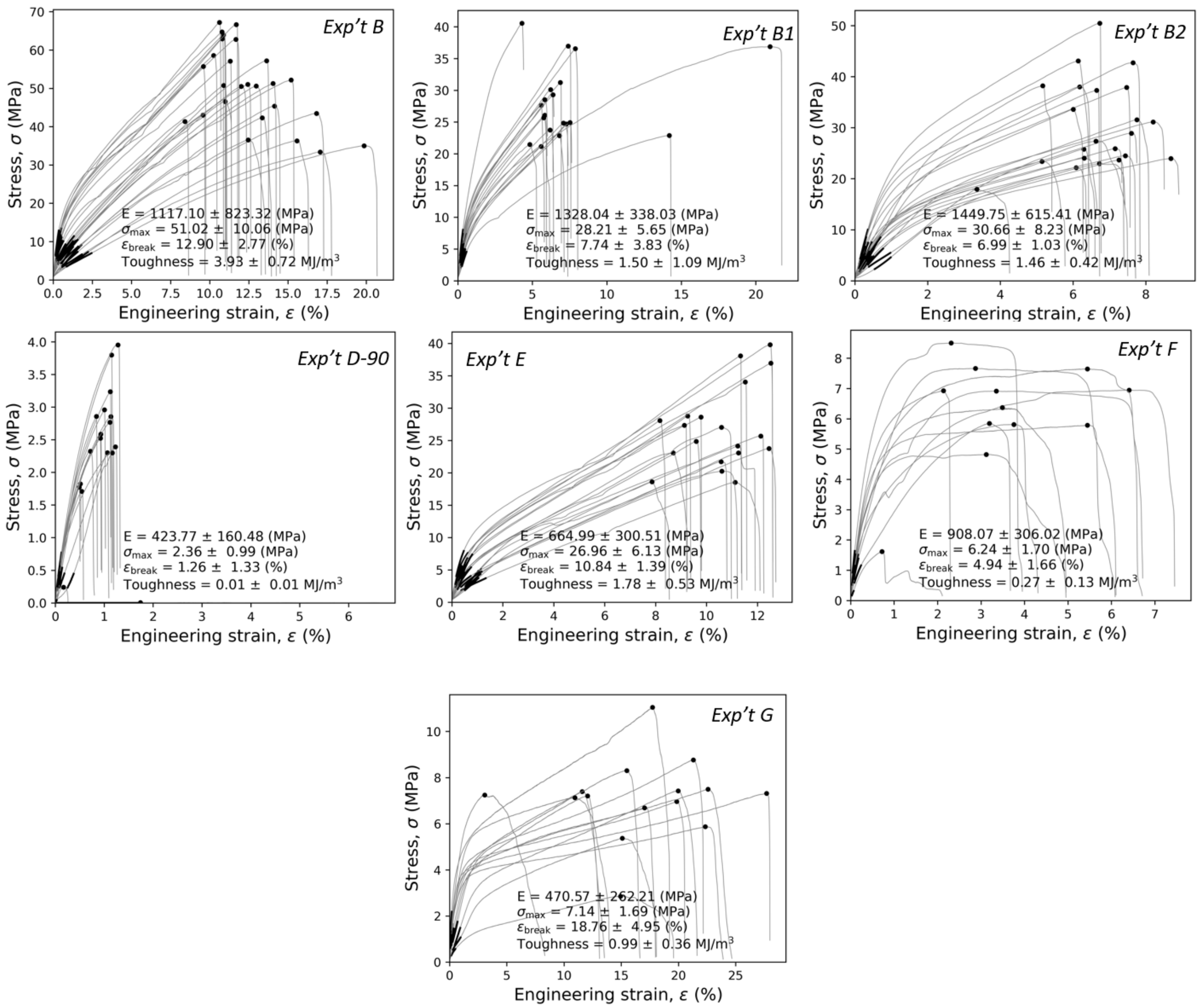


Figure 5.7: Stress-strain curves of experiment samples. Each curve represents the tensile test of an individual sample (outliers included). Scales differ from test to test to improve visibility of curve behavior.

5.5.5 *Machine learning model parameters*

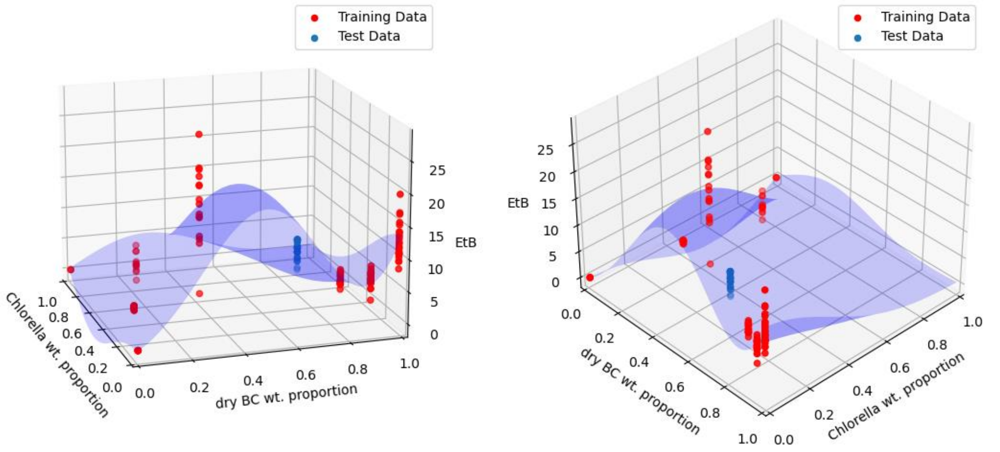
Once outliers were manually removed, data consisting of compositions and mechanical properties of individual test specimens were read as comma-separated values and grouped by composition. Values that were missing (in the case where the sample tested non-ideally) were ignored in the calculations of means and standard deviations of composition sets. Training and testing compositions were randomly selected and then individual values were standardized, in a statistical sense, using z-score normalization. By z-scoring values, the data is transformed and scaled about a mean with a standard deviation of 1. Some Gaussian Process models require data to be standardized, especially if the features have different scales, to prevent skewing of the model. The model presented hereafter did not require z-scoring of values, and thus unscored training data was passed into the model.

Once the model was trained, a normalized composition space was created. This space consisted of a grid bounded from [0 to 1], [0 to 1], and [-1 to 1] in the *x*, *y*, and *z* dimensions, respectively. The *z*-dimension, corresponding to the third compositional space, was defined as: 1 – (composition 1) – (composition 2). Defined as such, compositions 1 and 2 may vary while the third dimension is constrained by the sum of the other two compositions. Mathematically, this space is easy for the model to work in, but any composition sets that do not sum to 1 are unreal. More information can be found in the code in **Appendix A**.

5.5.6 Machine learning model output

The model may be represented in different forms. **Fig. 5.8** shows a 3D depiction with two compositions (BC and chlorella, in these cases) on the *x*-*y* plane and the respective mechanical property in the *z*-direction. Due to the constraints of the third composition, the value of the third composition can be calculated from the values of the other two compositions. These figures show that the model is indeed pinned to the average of the training data sets. That is, the light blue curve is the property prediction curve and it exactly predicts the average value at each training set composition. This data can also be flattened and represented as a 2D contour plot (**Fig. 5.9**). Real compositional values (i.e., composition 1 + composition 2 + composition 3 = 1) fall within the red dashed lines in the figure. Due to the full cube representation of the composition in the model, it is possible for the highest predicted values to lie outside of physically possible compositions, as in the case of toughness. For toughness, values outside the lines indicate that a two-component composite of a specific composition may have highest toughness. Additionally, the model predicts negative values of mechanical properties for a range of composition, which is unphysical but occurs due to the smoothness constraint imposed by the (RBF) kernel.

EtB



Toughness

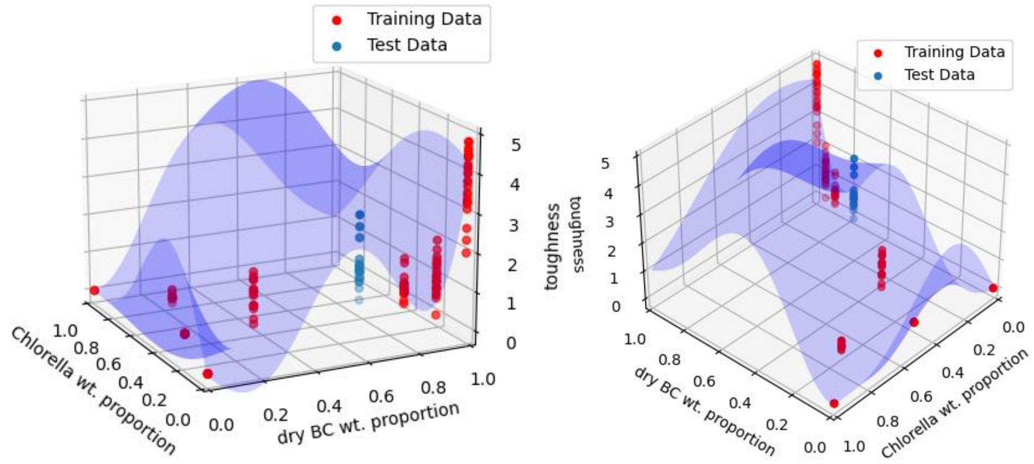


Figure 5.8: Predictive machine learning model EtB and Toughness curve representations. Red sets of dots indicate the training data that the model was trained on while the blue dots indicate test data that the model output was validated against. The units for toughness are MJ/m^3. The units for EtB are %. The grid representations are tilted differently only to improve curve visibility.

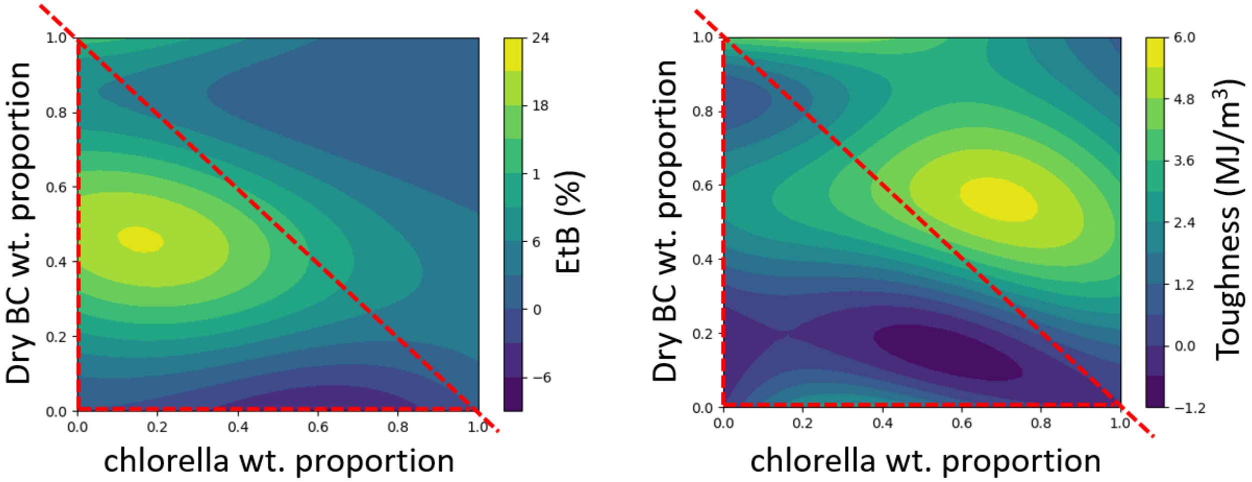


Figure 5.9: 2D contour plots depicting model predictions out of bounds. The areas bound by the red lines indicate real, possible compositions for two of the components. The composition of the third component is (1 – the sum of the other two components). Intersections on the red lines indicate 0 wt. proportion of the third component. The model is able to predict impossible negative values for properties and unreal compositions for composites due to the set-up of the training space. Extrema outside of physical ranges do not necessarily indicate a poor model, though the values may not be observed.

Negative values can be minimized by selecting more appropriate kernels for the data, obtaining more data points, and by changing the length scale of the model. **Fig. 5.10** illustrates the importance of appropriate length scale selection. While each predicted toughness curve in the figure passes through the given training data, the shapes of the curves differ. Larger numbers for length scales imply a relationship between more disparate data points and compositions while smaller numbers assume that only nearby compositions strongly influence the shape of the curve. At their limits, larger length scales would assume a relationship between all given compositions (that is, similar and dissimilar compositions simultaneously behave like each other despite potentially having different properties), while smaller length scales would only represent properties at the training data because the model is unable to extrapolate that the surrounding compositions to a training data point should behave similarly to the given data (assuming smoothness and connectedness). Therefore, it is imperative to select an appropriate length scale for the data to ensure accurate and fruitful predictions.

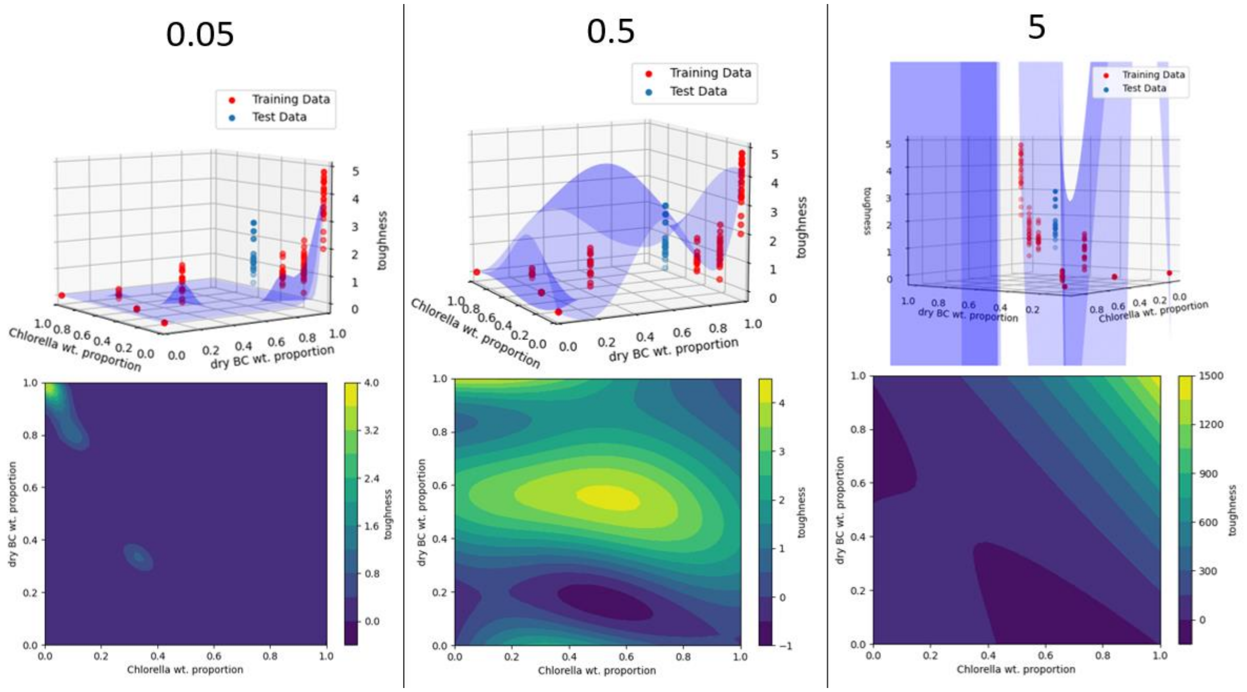


Figure 5.10: Effects of different length scales. The numbers above each column represent the length scale (unitless). By altering the length scales, both the magnitudes of the predicted property (toughness) and the compositions at which the properties manifest differ. In this case, 0.5 balances the influence of data points on its neighbors' properties while also allowing realistic extrapolations to dissimilar or untested compositional data.

There are methods that exist to optimize the length scale hyperparameter in the model. One of the methods tried here is to train a model, overlay a test data set, and then vary the length scales to find the length scale that passes through or near all of the training data means while also passing through the unseen test data. **Fig. 5.11** illustrates this method using the toughness plots shown in **Figs. 5.8** and **5.10**. In these plots, the red points indicate the training data set while the blue points indicate the test data used to validate against the model's predictive property curve. The mean toughness was found for the test data set (1.702 ± 0.554 MJ/m^3) and then the GP model was fit to the same RBF kernel, varying only the length scale in order to find the curve whose predicted toughness is similar to that of the test data at the test data's composition.

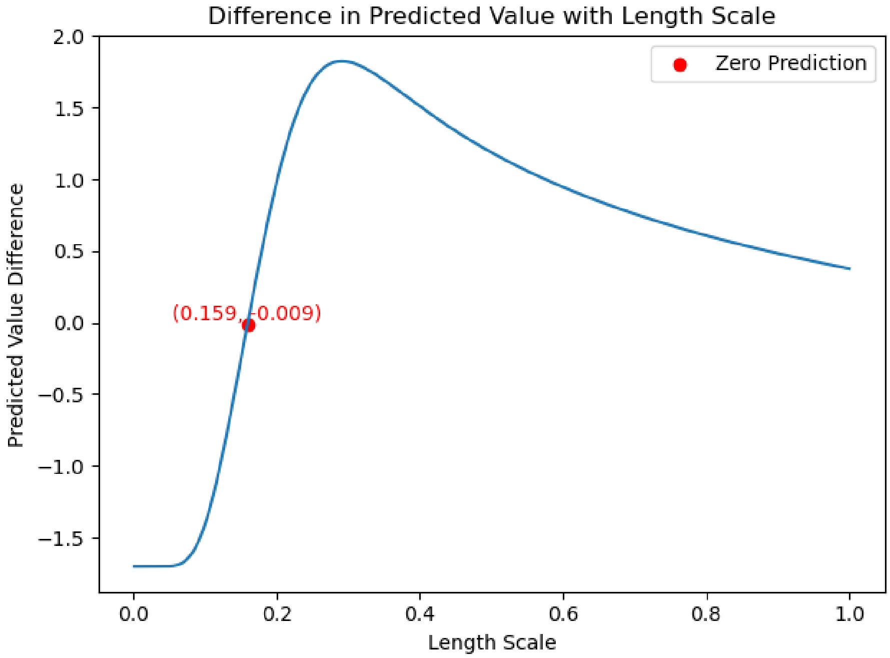


Figure 5.11: Difference in predicted value with length scale (toughness data). This curve was created by training the model on the same training data set, varying only the length scale. The model was asked to predict the output value at a particular composition that it had never seen before (the test data set). The curve was then outputted as the difference between the predicted value and the real value at that particular data point for length scales ranging from 0.001 to 1 in 0.001 increments.

Through varying the length scale of the model, we find that the length scale that most accurately describes the test data is 0.159, showing only a difference of 0.009 MJ/m^3 between the predicted and measured values at the measured values' composition.

One caveat with using this method is that passing through the test data is not necessarily an indication of a good fit. If the model is overfitted (in this case, due to an inappropriate length scale), the curve could pass through the test data but not generalize well to data at completed untested compositions (additional explanation of overfitting in the context of model complexity can be found in **Section 5.3.2 Machine learning terms and definitions**). Fortunately, some machine learning libraries come with optimization algorithms that can be used to optimize hyperparameters during training of the model. In this project, a second GP model was trained iteratively on the data (code for this algorithm can be found in **Appendix A**) to independently find an optimized length scale. In brief, the model is trained and the predicted toughness curve (output) is compared to the pinning point training data. The negative marginal log likelihood (NMLL), a function consisting of a term measuring how well the prediction matches the data and another term penalizing overly complex models, is iteratively minimized by changing the model parameters in such a way that the NMLL becomes closer to 0. By minimizing the NMLL, the least complex curve sufficient to fit the training data is produced. Here, after 250 iterations, the optimized length scale was reported by the model to be 0.196, close to the length scale of 0.159 obtained earlier from manually fitting the test data.

Another way to represent a 3-component system is by plotting on a ternary contour plot, shown in **Fig. 5.12**. The predicted model curve is flattened onto a 3-axes, two-dimensional plot. This plot has the benefit that only real compositions are present, so model predicted curves outside of real experimental conditions are not shown. **Table 5.3** summarizes model output and a comparison to the test data based on the optimized predictive curve fit. For both toughness and EtB, the model predicts higher values at the test data conditions than are actually observed. In EtB, there is a 39% difference while in toughness, there is a 42% difference. Despite the differences between the predictions and test data, the predictions are reasonable overall – the maximum mean values predicted across all compositions are 22.3% and 4.09 MJ/m^3. In comparison, the maximum toughness of any one sample was 4.80 MJ/m^3 (experiment B [0.5, 0.99, 0.5]) and the maximum

EtB was 27.7%. While the predictions are only means, they do not egregiously exceed actually observed values, suggesting a good model fit to the data.

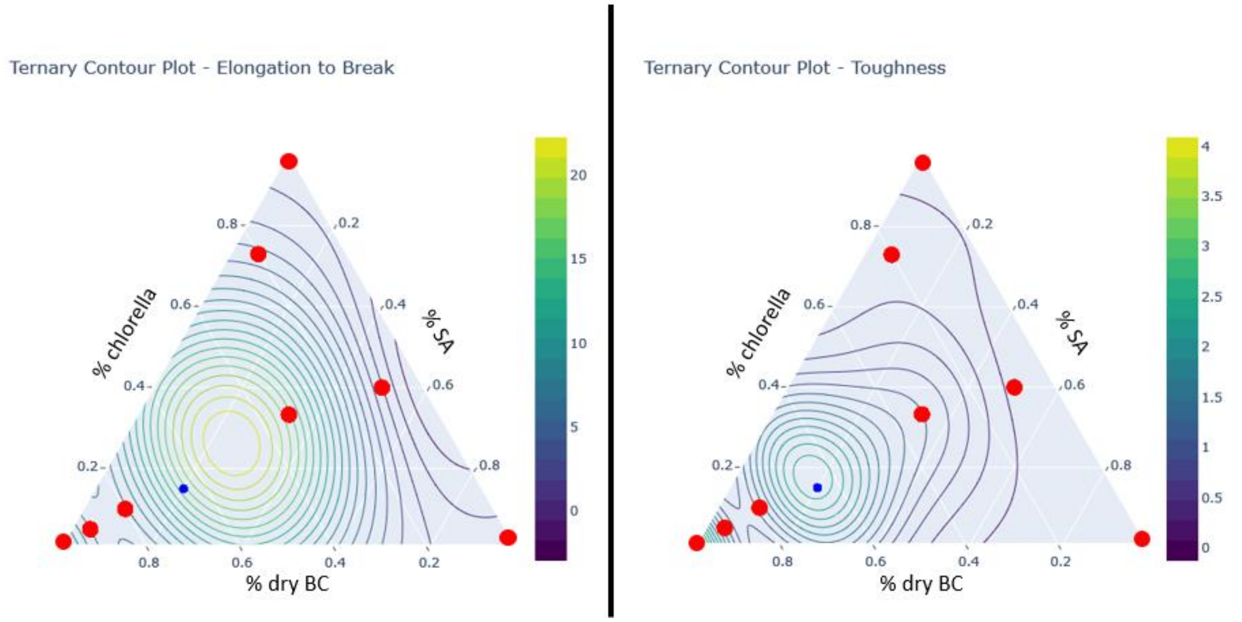


Figure 5.12: Ternary contour plots of Toughness and EtB, showing constraints to 3 compositions. Red dots indicate test data at their compositional intersections. The blue dot indicates the test data set (chlorella = 0.2, BC = 0.65, and SA = 0.20). Both plots use a model with their optimized length scales.

Table 5.3: Comparisons between model predictions and real test data.
(Key: Tgh = toughness; EtB = elongation at max stress before failure)

	Length scale	*Max predicted value (MPV)*	*Composition at MPV (C, BC, SA)*	*Second max value (SMV)*	*Composition at SMV (C, BC, SA)*
Model EtB	*0.299*	*22.3%*	*(0.27, 0.49, 0.24)*	*22.1%*	*(0.24, 0.51, 0.24)*
Model Tgh	*0.196*	*4.09 MJ/m^3*	*(0.0, 1.0, 0.0)*	*2.63 MJ/m^3*	*(0.18, 0.65, 0.16)*

	Test data composition (C, BC, SA)	*Test data Property (n = 17)*	*Predicted Property at test data composition*
EtB	*(0.15, 0.65, 0.20)*	*10.61 ± 1.46 %*	*15.73 %*
Tgh	*(0.15, 0.65, 0.20)*	*1.70 ± 0.55 MJ/m^3*	*2.60 MJ/m^3*

The composition predicted to yield the most extensible samples is [chlorella = 0.27, BC = 0.49, SA = 0.24]. This composition is most similar to that of experiments G and E, which also yielded among the most extensible samples. In the case of toughness, pure BC was predicted to yield the best properties at 4.09 MJ/m^3, while the composition of [chlorella = 0.18, BC = 0.65, SA = 0.16] was expected to yield the second highest set of values at 2.63 MJ/m^3. Interestingly, the predicted toughness for pure BC films is on the order of the strongest processed BC obtained in **Chapter 4**, whose average value across all processing conditions was 2.48 ± 1.03 MJ/m^3 (with the toughest samples having 4.46 ± 1.15 MJ/m^3), again corroborating the realistic prediction of the ML model for BC materials. For toughness, the B series of experiments compares closest in composition while the second highest set of toughnesses belongs to a composition similar to that of experiment E again. Because toughness is related to extensibility insofar as toughness increases with extensibility, it is not unrealistic to see that experiment E yields both extensible and tough samples. The more interesting aspect is that these compositions are dominant in the BC component. Recalling from the results of **Chapter 4**, BC fibrils appeared to be the primary load-bearing element in the composite system. For a composite system that contains BC whose network was disrupted via processing and potentially distributed or broken up by the presence of chlorella and stearic acid components, does this behavior still hold?

5.5.7 SEM analysis and structure-property link

To resolve this question, SEM analysis was conducted on samples of experiment G (Composition [0.33, 0.33, 0.34]). Images were taken at different angles of the fracture surface of tensile specimens of experiment G and are presented in **Fig. 5.13**. The top view shows that the composite is uniformly mixed and the blended components form a cohesive matrix. We did not observe any phase separation or otherwise distinct arrangement of individual components. Away from the fracture surface, BC fibers are well-impregnated in and indistinguishable from the bulk matrix. In comparison, the fracture surface reveals that BC fibers are pulled out of the bulk composite matrix. Furthermore, fiber bundles show varying degrees of defibrillation; both micro-sized bundles and individual nanofibers are present. Fiber pull-out and defibrillation are toughening mechanisms, allowing the composite to sustain substantial deformation prior to failure and supporting the observed high elongation before failure. Toughness on par with BC-dominant composites (E: 65 wt.%; B1: 80 wt.%, B2: 90 wt.%) suggests that BC alone is not the main factor in the toughening of the material.

Raw biomatter cells result in brittle and low strength materials. This statement is supported by the results in **Chapter 3** that showed that spirulina-only materials were untestable due to their brittleness and required a strengthening additive (α-cellulose) to be able to measure properties. In this project, the samples with the highest proportion of chlorella, experiments A (99 wt.% - too brittle to be tested), F (73 wt.%), and D-90 (40 wt.%), showed the lowest strength and elongation to break. In contrast, BC is the strongest component in this system and thus the component that we want the stress to be transferred to. Stress transference in a composite is optimized when there is good interfacial adhesion between the matrix and the filler (BC fibers, in this case). Thus, an explanation as to why experiment G, containing only 33 wt.% BC, shows properties similar to or exceeding that of greater BC-proportion-containing samples is because of the role of the other components in the composite.

As evidenced by experiment D-90 (50 wt.% stearic acid) that showed poor mechanical properties, stearic acid alone is not a primary contributor to the strengthening or toughening of a composite. However, in conjunction with chlorella, it may form a matrix capable of binding cellulose. While chlorella is hydrophilic and may interact better with BC fibrils, stearic acid is smaller (i.e., on the molecular scale) and may be more evenly dispersed. We hypothesize that in this ternary system, stearic acid functions as a plasticizer by improving BC fiber sliding over a larger scale despite being mostly hydrophobic. Due to the relatively equal proportions of the non-BC components in the E, B1, and B2 experiments, further investigation will be required to elucidate how and how much stearic acid and chlorella play a role in facilitating BC fiber rearrangement or sliding, leading to potentially higher extensibility as remarked in **Chapter 4**. However, the results suggest that a critical amount of either or both are needed for their hypothesized roles in improving the mechanical properties of the composite, which is consistent with our machine learning model and the bioplastics literature.[142]

We conclude that in our ternary system of chlorella cells, BC fibers, and stearic acid molecules, stearic acid functions as a plasticizer and chlorella, in small amounts, forms a cohesive matrix to embed the BC fibers. BC confers strength and substantial energy dissipation mechanisms prior to failure, and stearic acid working in conjunction with chlorella helps form a compatible matrix that allows stresses to be effectively transferred to the load-bearing BC. When the amount of chlorella is greatly increased or BC is greatly decreased, we observe drastic losses in strength, EtB, and toughness. When stearic acid is increased in presence of chlorella, we see an increase in EtB, but not toughness as the strength reductions caused by chlorella do not allow overall improvement in toughness.

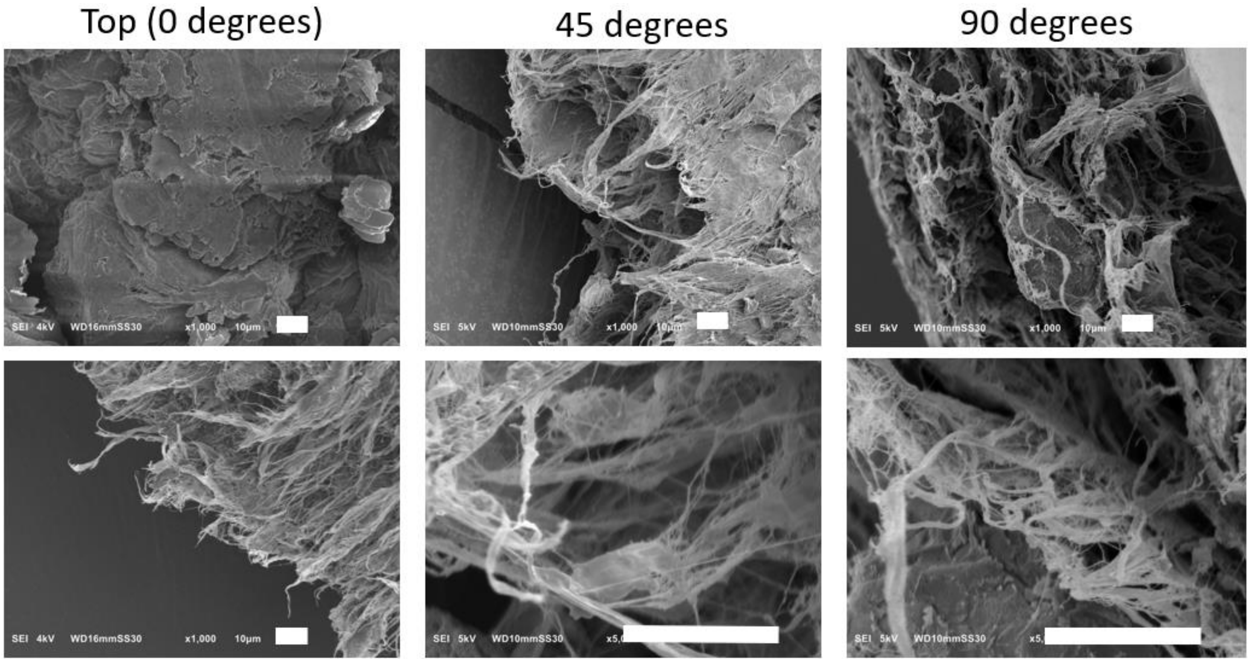


Figure 5.13: SEM images of experiment G composite film structures at fracture surfaces. The angle measures at the top of each column of images indicate only differences between SEM stub mounting angles of the specimens. Otherwise, each sample was tensile tested to failure under the same conditions. White bars indicate 10 μm length.

5.6 Summary and Conclusions

In this final chapter, composite films of chlorella biomass/cells, bacterial cellulose fibers, and stearic acid molecules were prepared by a blending and casting procedure. An initial set of dry mass compositions were selected as input data and their corresponding mechanical properties were used as output data; both data were used as the basis training data set for the machine learning model. Nearly single component composites of chlorella and stearic acid, respectively, were too brittle to be tested while a nearly single component BC film was testable. Intermediate compositions were all castable and testable, and thus were successfully used as training data sets.

A short series of experiments was conducted whereby different absolute amounts of solvent were combined with the same dry mass compositions. At excessive solvent amounts, the films were practically unable to be casted and led to untestable films. At lower solvent amounts, films were produced with different drying characteristics such as warping behavior. Tensile testing results showed that there was a 1.5- to 3-fold difference in properties in the resultant films. Further investigation is required to elucidate the mechanisms by which absolute solvent amount affects the properties of films. These experiments were used to establish the mixing protocol such that a 1.6 ethanol/water mass ratio would be applied to each composite in the wet state to facilitate blending and ensure dissolution of stearic acid.

Viscosity measurements were made of the pre-casted composite mixtures. All solutions tested showed shear thinning rheological behavior. There was no apparent relationship between the concentration of specific components and their viscosity measurements. In the future, this data may be used as additional, potentially informative, features for the machine learning model.

A brief glance at the tensile testing data (mechanical properties) revealed the presence of outliers, which were removed for the purpose of feeding into the machine learning algorithm in order to mitigate or minimize erroneous predictions. A 3D rectangular space was created as a mathematical representation of the ternary composition composites, such that each 3-coordinate point was associated with a single mechanical property of interest to optimize. As example systems, toughness and elongation at max stress were used separately as the mechanical properties of interest. The model was able to successfully learn and extrapolate from the given training sets; it produced mechanical property predictions at all compositions in the virtual space, including at unreal compositions (i.e, compositions whose proportions sum to 1 and cannot be synthesized but are represented in this space due to the set-up of the model).

In the model (hyper)parameters, length scale was shown to be critical for the production of an accurate model. Too small of a length scale would result in a model that could not extrapolate to unseen/untested compositions given a small training set. Too large of a length scale would result in an overfitted and complex model that produced intuitively unrealistic predictions based on the observations. Two ways were demonstrated to optimize the model hyperparameters, including the use of a separate model whose hyperparameter optimization algorithm could be manually adjusted to suggest the most appropriate length scale and hence best-fitting predictive curve.

Ternary contour plots were used to showcase the differences between our resulting predictive model with optimized length scale and a testing composition set that was unseen by the model. The comparison showed that the model overestimated the toughness (2.6 MJ/m^3 vs. 1.7 MJ/m^3) and extensibility (15.7% vs. 10.6%) at the test composition by about 40%. Furthermore, the model also predicted peak values could be obtained at intermediate compositions, implying compatibility between the composite components.

A final SEM analysis on a (33 wt.%, 33 wt.%, 34 wt.%) composite corroborated BC as the load-bearing component in the system. At this ratio, there was no apparent phase separation or distinct individual components, suggesting a uniformly mixed and compatible matrix. We noted that while raw biomatter (chlorella/spirulina) usually results in poor mechanical properties, its combination with stearic acid seemed to mitigate some of the negative effects given that a sufficient amount of BC was present to strengthen the composite. For this sample, an increased elongation to break was observed over any other composition, but it presented low strength given that it contains a relatively small amount of BC. This tradeoff between properties, rather than a complete breakdown in properties due to increased biomatter/decreased structural components, suggests a constructive relationship between the components. We hypothesized that in this ternary system, stearic acid functions as a plasticizer when present in sufficient amounts with chlorella strengthened with BC. Further investigation will be necessary to establish how these components work together to form composites that are predicted to be optimized by an intermediate composition rather than an extreme composition.

In conclusion, a machine learning model was trained on a small set of data and predicted output that was reasonable overall (i.e., within previously observed values for the individual components), overestimated test data by a small amount (40% difference), and predicted optimal compositions that are also plausible based on mechanistic reasoning. As a tool, machine learning has the potential to cut down on time and cost of unproductive experimentation. We demonstrated that even a relatively small set of data (8 experiments) could successfully be used to gain information about the relationships between components in a system and to guide selection of subsequent experiments more efficiently and rationally.

6. Summary and Outlook

This dissertation work has shown three ways to design composites of biomass-based materials (*biomatter*). The first method is the traditional approach of trial and error. In this approach, subsequent iterations are based on reasonable interpretations of data and intuition based on prior experience of chemically similar systems. The second method is a statistical, design of experiments approach. A prescribed set of experiments of a specific format are conducted and statistical tools exist to analyze the results and gain as much information as possible from a minimal amount of experiments. The third method is a machine learning approach. A model is developed that draws statistical inferences from underlying patterns in the data. The patterns may be arbitrary or non-analytical, but the model's precision and accuracy increase as the amount of data or experimental information provided increases, given that there is an actual pattern or relationship between experimental variables.

In this work, processing and compositional parameters were varied and studied. The first and second chapters, using the traditional and design of experiments approaches, respectively, studied both processing and composition variation. The third chapter's focus was explicitly compositional variation. We showed that these methods of design are generalizable: from abstract processing parameters to specific, sometimes complex, materials composition. The results of each of these approaches are equivalent, in theory, while only the process changes. The traditional approach allows for the fastest time to start, having the simplest theoretical considerations and requiring minimal planning. However, there is a tradeoff with analysis time – the often inefficient process of the traditional approach makes searching a parameter or variable space for optimal properties difficult due to the lack of systems insight that is provided from the data. In this regard, design of experiments offers the most robust information as far as statistics allows. However, it has the disadvantage of the user having to plan a very specific, prescribed set of experiments in order to make use of the statistical information. Finally, the machine learning approach allows for a balance of experimental start-up time (in that an experiment can be started with any set of parameters and fed as data into a model) and outputted insight given that an underlying relationship between variables exist. However, the drawback of this method, in particular the simple regression-based Gaussian Process modeling that we used, is that its inferences are a black box – when insights or predictions are correct but unintuitive, the model itself cannot be used to explain the results mechanistically or theoretically. This issue can be overcome by the use of more complex machine learning architectures, such as neural networks. Ongoing research has shown that neural networks can integrate physical theory or physical constraints in their loss functions and inference algorithms to favor solutions that follow physical laws and penalize or disfavor solutions that do not.[143,144] These approaches have been called physics-informed[143] or theory-guided[144] machine learning.

One of the biggest hurdles with rational materials design in systems made of complex materials like biomass is understanding the underlying structure/component-property relationships that govern their interactions with other components. Machine learning has the potential to allow for relatively quick insights from relatively small data sets, a common feature of some fields of materials science in which conducting experiments may be cost- or time-prohibitive. Rational materials design shares a parallel challenge with machine learning models: to design materials rationally, their governing physical and chemical principles must be known; to use machine learning effectively, informative features must be provided to the model. However, to know which features are informative (and can be used to explain the outputs) is to know, to a degree, how a material works at a fundamental level. Part of the difficulty with feature selection in machine learning is understanding which features are informative. Without understanding this, a machine learning approach reduces to a traditional approach – parameters are varied and outputs are measured but there are no indications or suggestions of what would work better and for what reason. Design of experiments and machine learning are useful tools for understanding the relationships between a large number of simultaneous variables, but they are not sufficient or appropriate tools for explaining why. To design materials rationally, these methods must be used in conjunction with analytical techniques to streamline and speed up the experimentation and inference-drawing process.

References

(1) Rochman, C. M.; Browne, M. A.; Halpern, B. S.; Hentschel, B. T.; Hoh, E.; Karapanagioti, H. K.; Rios-Mendoza, L. M.; Takada, H.; Teh, S.; Thompson, R. C. Classify Plastic Waste as Hazardous. *Nature* **2013**, *494* (7436), 169–171. https://doi.org/10.1038/494169a.

(2) Mohanty, A. K.; Vivekanandhan, S.; Pin, J.-M.; Misra, M. Composites from Renewable and Sustainable Resources: Challenges and Innovations. *Science* **2018**, *362* (6414), 536–542. https://doi.org/10.1126/science.aat9072.

(3) Reddy, M. M.; Vivekanandhan, S.; Misra, M.; Bhatia, S. K.; Mohanty, A. K. Biobased Plastics and Bionanocomposites: Current Status and Future Opportunities. *Prog. Polym. Sci.* **2013**, *38* (10), 1653–1689. https://doi.org/10.1016/j.progpolymsci.2013.05.006.

(4) Schneiderman, D. K.; Hillmyer, M. A. 50th Anniversary Perspective: There Is a Great Future in Sustainable Polymers. *Macromolecules* **2017**, *50* (10), 3733–3749. https://doi.org/10.1021/acs.macromol.7b00293.

(5) Haque, F. M.; Ishibashi, J. S. A.; Lidston, C. A. L.; Shao, H.; Bates, F. S.; Chang, A. B.; Coates, G. W.; Cramer, C. J.; Dauenhauer, P. J.; Dichtel, W. R.; Ellison, C. J.; Gormong, E. A.; Hamachi, L. S.; Hoye, T. R.; Jin, M.; Kalow, J. A.; Kim, H. J.; Kumar, G.; LaSalle, C. J.; Liffland, S.; Lipinski, B. M.; Pang, Y.; Parveen, R.; Peng, X.; Popowski, Y.; Prebihalo, E. A.; Reddi, Y.; Reineke, T. M.; Sheppard, D. T.; Swartz, J. L.; Tolman, W. B.; Vlaisavljevich, B.; Wissinger, J.; Xu, S.; Hillmyer, M. A. Defining the Macromolecules of Tomorrow through Synergistic Sustainable Polymer Research. *Chem. Rev.* **2022**, *122* (6), 6322–6373. https://doi.org/10.1021/acs.chemrev.1c00173.

(6) Hillmyer, M. A.; Tolman, W. B. Aliphatic Polyester Block Polymers: Renewable, Degradable, and Sustainable. *Acc. Chem. Res.* **2014**, *47* (8), 2390–2396. https://doi.org/10.1021/ar500121d.

(7) Garcia, J. M.; Robertson, M. L. The Future of Plastics Recycling. *Science* **2017**, *358* (6365), 870–872. https://doi.org/10.1126/science.aaq0324.

(8) Lu, H.; Diaz, D. J.; Czarnecki, N. J.; Zhu, C.; Kim, W.; Shroff, R.; Acosta, D. J.; Alexander, B. R.; Cole, H. O.; Zhang, Y.; Lynd, N. A.; Ellington, A. D.; Alper, H. S. Machine Learning-Aided Engineering of Hydrolases for PET Depolymerization. *Nature* **2022**, *604* (7907), 662–667. https://doi.org/10.1038/s41586-022-04599-z.

(9) Fortman, D. J.; Brutman, J. P.; De Hoe, G. X.; Snyder, R. L.; Dichtel, W. R.; Hillmyer, M. A. Approaches to Sustainable and Continually Recyclable Cross-Linked Polymers. *ACS Sustain. Chem. Eng.* **2018**, *6* (9), 11145–11159. https://doi.org/10.1021/acssuschemeng.8b02355.

(10) Liptow, C.; Tillman, A.-M. A Comparative Life Cycle Assessment Study of Polyethylene Based on Sugarcane and Crude Oil. Rochester, NY June 1, 2012. https://doi.org/10.1111/j.1530-9290.2011.00405.x.

(11) Hillmyer, M. A. The Promise of Plastics from Plants. *Science* **2017**, *358* (6365), 868–870. https://doi.org/10.1126/science.aao6711.

(12) Moon, R. J.; Martini, A.; Nairn, J.; Simonsen, J.; Youngblood, J. Cellulose Nanomaterials Review: Structure, Properties and Nanocomposites. *Chem. Soc. Rev.* **2011**, *40* (7), 3941–3994. https://doi.org/10.1039/c0cs00108b.

(13) Li, T.; Chen, C.; Brozena, A. H.; Zhu, J.; Xu, L.; Driemeier, C.; Dai, J.; Rojas, O. J.; Isogai, A.; Wågberg, L.; others. Developing Fibrillated Cellulose as a Sustainable Technological Material. *Nature* **2021**, *590* (7844), 47–56.

(14) Meng, Q.; Wang, T. J. Mechanics of Strong and Tough Cellulose Nanopaper. *Appl. Mech. Rev.* **2019**, *71* (4).

(15) Wang, X.; Xia, Q.; Jing, S.; Li, C.; Chen, Q.; Chen, B.; Pang, Z.; Jiang, B.; Gan, W.; Chen, G.; Cui, M.; Hu, L.; Li, T. Strong, Hydrostable, and Degradable Straws Based on Cellulose-Lignin Reinforced Composites. *Small* **2021**, *17* (18), 2008011. https://doi.org/10.1002/smll.202008011.

(16) Xia, Q.; Chen, C.; Yao, Y.; Li, J.; He, S.; Zhou, Y.; Li, T.; Pan, X.; Yao, Y.; Hu, L. A Strong, Biodegradable and Recyclable Lignocellulosic Bioplastic. *Nat. Sustain.* **2021**, *4* (7), 627–635. https://doi.org/10.1038/s41893-021-00702-w.

(17) Chen, C.; Kuang, Y.; Zhu, S.; Burgert, I.; Keplinger, T.; Gong, A.; Li, T.; Berglund, L.; Eichhorn, S. J.; Hu, L. Structure–Property–Function Relationships of Natural and Engineered Wood. *Nat. Rev. Mater.* **2020**, *5* (9), 642–666. https://doi.org/10.1038/s41578-020-0195-z.

(18) Foster, E. J.; Moon, R. J.; Agarwal, U. P.; Bortner, M. J.; Bras, J.; Camarero-Espinosa, S.; Chan, K. J.; Clift, M. J.; Cranston, E. D.; Eichhorn, S. J.; others. Current Characterization Methods for Cellulose Nanomaterials. *Chem. Soc. Rev.* **2018**, *47* (8), 2609–2679.

(19) Zhu, H.; Zhu, S.; Jia, Z.; Parvinian, S.; Li, Y.; Vaaland, O.; Hu, L.; Li, T. Anomalous Scaling Law of Strength and Toughness of Cellulose Nanopaper. *Proc. Natl. Acad. Sci.* **2015**, *112* (29), 8971–8976. https://doi.org/10.1073/pnas.1502870112.

(20) Li, J.; Chen, C.; Zhu, J. Y.; Ragauskas, A. J.; Hu, L. In Situ Wood Delignification toward Sustainable Applications. *Acc. Mater. Res.* **2021**, *2* (8), 606–620. https://doi.org/10.1021/accountsmr.1c00075.

(21) Mao, R.; Goutianos, S.; Tu, W.; Meng, N.; Yang, G.; Berglund, L. A.; Peijs, T. Comparison of Fracture Properties of Cellulose Nanopaper, Printing Paper and Buckypaper. *J. Mater. Sci.* **2017**, *52* (16), 9508–9519.

(22) Wang, S.; Jiang, F.; Xu, X.; Kuang, Y.; Fu, K.; Hitz, E.; Hu, L. Super-Strong, Super-Stiff Macrofibers with Aligned, Long Bacterial Cellulose Nanofibers. *Adv. Mater.* **2017**, *29* (35), 1702498. https://doi.org/10.1002/adma.201702498.

(23) Yeap, R. Y. The Potential of Lignin to Increase the Hydrophobicity of Micro/Nanofibrillated Cellulose (MNFC). PhD Thesis, University of British Columbia, 2020.

(24) Jiang, B.; Chen, C.; Liang, Z.; He, S.; Kuang, Y.; Song, J.; Mi, R.; Chen, G.; Jiao, M.; Hu, L. Lignin as a Wood-Inspired Binder Enabled Strong, Water Stable, and Biodegradable Paper for Plastic Replacement. *Adv. Funct. Mater.* **2020**, *30* (4), 1906307. https://doi.org/10.1002/adfm.201906307.

(25) Farooq, M.; Zou, T.; Riviere, G.; Sipponen, M. H.; Österberg, M. Strong, Ductile, and Waterproof Cellulose Nanofibril Composite Films with Colloidal Lignin Particles. *Biomacromolecules* **2019**, *20* (2), 693–704. https://doi.org/10.1021/acs.biomac.8b01364.

(26) Song, J.; Chen, C.; Zhu, S.; Zhu, M.; Dai, J.; Ray, U.; Li, Y.; Kuang, Y.; Li, Y.; Quispe, N.; Yao, Y.; Gong, A.; Leiste, U. H.; Bruck, H. A.; Zhu, J. Y.; Vellore, A.; Li, H.; Minus, M. L.; Jia, Z.; Martini, A.; Li, T.; Hu, L. Processing Bulk Natural Wood into a High-Performance Structural Material. *Nature* **2018**, *554* (7691), 224–228. https://doi.org/10.1038/nature25476.

(27) Xiao, S.; Chen, C.; Xia, Q.; Liu, Y.; Yao, Y.; Chen, Q.; Hartsfield, M.; Brozena, A.; Tu, K.; Eichhorn, S. J.; Yao, Y.; Li, J.; Gan, W.; Shi, S. Q.; Yang, V. W.; Lo Ricco, M.; Zhu, J. Y.; Burgert, I.; Luo, A.; Li, T.; Hu, L. Lightweight, Strong, Moldable Wood via Cell Wall Engineering as a Sustainable Structural Material. *Science* **2021**, *374* (6566), 465–471. https://doi.org/10.1126/science.abg9556.

(28) Otsuki, T.; Zhang, F.; Kabeya, H.; Hirotsu, T. Synthesis and Tensile Properties of a Novel Composite of Chlorella and Polyethylene. *J. Appl. Polym. Sci.* **2004**, *92* (2), 812–816. https://doi.org/10.1002/app.13650.

(29) Zhu, N.; Ye, M.; Shi, D.; Chen, M. Reactive Compatibilization of Biodegradable Poly (Butylene Succinate)/Spirulina Microalgae Composites. *Macromol. Res.* **2017**, *25* (2), 165–171.

(30) Chiellini, E.; Cinelli, P.; Ilieva, V. I.; Martera, M. Biodegradable Thermoplastic Composites Based on Polyvinyl Alcohol and Algae. *Biomacromolecules* **2008**, *9* (3), 1007–1013. https://doi.org/10.1021/bm701041e.

(31) Hassan, M. M.; Mueller, M.; Wagners, M. H. Exploratory Study on Seaweed as Novel Filler in Polypropylene Composite. *J. Appl. Polym. Sci.* **2008**, *109* (2), 1242–1247. https://doi.org/10.1002/app.28287.

(32) Bulota, M.; Budtova, T. PLA/Algae Composites: Morphology and Mechanical Properties. *Compos. Part Appl. Sci. Manuf.* **2015**, *73*, 109–115. https://doi.org/10.1016/j.compositesa.2015.03.001.

(33) Devadas, V. V.; Khoo, K. S.; Chia, W. Y.; Chew, K. W.; Munawaroh, H. S. H.; Lam, M.-K.; Lim, J.-W.; Ho, Y.-C.; Lee, K. T.; Show, P. L. Algae Biopolymer towards Sustainable Circular Economy. *Bioresour. Technol.* **2021**, *325*, 124702. https://doi.org/10.1016/j.biortech.2021.124702.

(34) Onen Cinar, S.; Chong, Z. K.; Kucuker, M. A.; Wieczorek, N.; Cengiz, U.; Kuchta, K. Bioplastic Production from Microalgae: A Review. *Int. J. Environ. Res. Public. Health* **2020**, *17* (11). https://doi.org/10.3390/ijerph17113842.

(35) Zhang, C.; Show, P.-L.; Ho, S.-H. Progress and Perspective on Algal Plastics – A Critical Review. *Bioresour. Technol.* **2019**, *289*, 121700. https://doi.org/10.1016/j.biortech.2019.121700.

(36) Das, A. A. K.; Bovill, J.; Ayesh, M.; Stoyanov, S. D.; Paunov, V. N. Fabrication of Living Soft Matter by Symbiotic Growth of Unicellular Microorganisms. *J Mater Chem B* **2016**, *4* (21), 3685–3694. https://doi.org/10.1039/C5TB02489G.

(37) Zeller, M. A.; Hunt, R.; Jones, A.; Sharma, S. Bioplastics and Their Thermoplastic Blends from Spirulina and Chlorella Microalgae. *J. Appl. Polym. Sci.* **2013**, *130* (5), 3263–3275. https://doi.org/10.1002/app.39559.

(38) Mathiot, C.; Ponge, P.; Gallard, B.; Sassi, J.-F.; Delrue, F.; Moigne, N. L. Microalgae Starch-Based Bioplastics: Screening of Ten Strains and Plasticization of Unfractionated Microalgae by Extrusion. *Carbohydr. Polym.* **2019**, *208*, 142–151. https://doi.org/10.1016/j.carbpol.2018.12.057.

(39) Fredricks, J. L.; Iyer, H.; McDonald, R.; Hsu, J.; Jimenez, A. M.; Roumeli, E. Spirulina-Based Composites for 3D-Printing. *J. Polym. Sci.* **2021**, *59* (22), 2878–2894. https://doi.org/10.1002/pol.20210683.

(40) Roumeli, E.; Hendrickx, R.; Bonanomi, L.; Vashisth, A.; Rinaldi, K.; Daraio, C. Biological Matrix Composites from Cultured Plant Cells. *Proc. Natl. Acad. Sci.* **2022**, *In Press*.

(41) Noreen, A.; Zia, K. M.; Zuber, M.; Ali, M.; Mujahid, M. A Critical Review of Algal Biomass: A Versatile Platform of Bio-Based Polyesters from Renewable Resources. *Int. J. Biol. Macromol.* **2016**, *86*, 937–949. https://doi.org/10.1016/j.ijbiomac.2016.01.067.

(42) Pellis, A.; Malinconico, M.; Guarneri, A.; Gardossi, L. Renewable Polymers and Plastics: Performance beyond the Green. *New Biotechnol.* **2021**, *60*, 146–158. https://doi.org/10.1016/j.nbt.2020.10.003.

(43) Ling, S.; Chen, W.; Fan, Y.; Zheng, K.; Jin, K.; Yu, H.; Buehler, M. J.; Kaplan, D. L. Biopolymer Nanofibrils: Structure, Modeling, Preparation, and Applications. *Prog. Polym. Sci.* **2018**, *85*, 1–56. https://doi.org/10.1016/j.progpolymsci.2018.06.004.

(44) Michalak, I.; Mironiuk, M.; Godlewska, K.; Trynda, J.; Marycz, K. Arthrospira (Spirulina) Platensis: An Effective Biosorbent for Nutrients. *Process Biochem.* **2020**, *88*, 129–137. https://doi.org/10.1016/j.procbio.2019.10.004.

(45) Pollak, M.; Kocisko, M.; Basistova, A.; Hlavata, S. PRODUCTION OF FIBER AS AN INPUT MATERIAL FOR THE 3D PRINTING PROCESS. *MM Sci. J.* **2021**, *2021* (2), 4414–4419. https://doi.org/10.17973/MMSJ.2021_6_2021031.

(46) Zhu, C.; Li, T.; Mohideen, M. M.; Hu, P.; Gupta, R.; Ramakrishna, S.; Liu, Y. Realization of Circular Economy of 3D Printed Plastics: A Review. *Polymers* **2021**, *13* (5), 744. https://doi.org/10.3390/polym13050744.

(47) Faruk, O.; Bledzki, A. K.; Fink, H.-P.; Sain, M. Biocomposites Reinforced with Natural Fibers: 2000–2010. *Prog. Polym. Sci.* **2012**, *37* (11), 1552–1596.

(48) Li, M.; Pu, Y.; Thomas, V. M.; Yoo, C. G.; Ozcan, S.; Deng, Y.; Nelson, K.; Ragauskas, A. J. Recent Advancements of Plant-Based Natural Fiber–Reinforced Composites and Their Applications. *Compos. Part B Eng.* **2020**, *200*, 108254. https://doi.org/10.1016/j.compositesb.2020.108254.

(49) Mazzanti, V.; Malagutti, L.; Mollica, F. FDM 3D Printing of Polymers Containing Natural Fillers: A Review of Their Mechanical Properties. *Polymers* **2019**, *11* (7), 1094.

(50) Jiang, Y.; Raney, J. R. 3D Printing of Amylopectin-Based Natural Fiber Composites. *Adv. Mater. Technol.* **2019**, *4* (11), 1900521.

(51) Babu, A.; Kumaresan, G.; Raj, V. A. A.; Velraj, R. Review of Leaf Drying: Mechanism and Influencing Parameters, Drying Methods, Nutrient Preservation, and Mathematical Models. *Renew. Sustain. Energy Rev.* **2018**, *90*, 536–556.

(52) Karakurt, I.; Lin, L. 3D Printing Technologies: Techniques, Materials, and Post-Processing. *Curr. Opin. Chem. Eng.* **2020**, *28*, 134–143. https://doi.org/10.1016/j.coche.2020.04.001.

(53) Håkansson, K. M.; Henriksson, I. C.; de la Peña Vázquez, C.; Kuzmenko, V.; Markstedt, K.; Enoksson, P.; Gatenholm, P. Solidification of 3D Printed Nanofibril Hydrogels into Functional 3D Cellulose Structures. *Adv. Mater. Technol.* **2016**, *1* (7), 1600096.

(54) Lefevere, J.; Protasova, L.; Mullens, S.; Meynen, V. 3D-Printing of Hierarchical Porous ZSM-5: The Importance of the Binder System. *Mater. Des.* **2017**, *134*, 331–341.

(55) Shahbazi, M.; Jäger, H. Current Status in the Utilization of Biobased Polymers for 3D Printing Process: A Systematic Review of the Materials, Processes, and Challenges. *ACS Appl. Bio Mater.* **2020**, *4* (1), 325–369.

(56) Dizon, J. R. C.; Gache, C. C. L.; Cascolan, H. M. S.; Cancino, L. T.; Advincula, R. C. Post-Processing of 3D-Printed Polymers. *Technologies* **2021**, *9* (3), 61.

(57) Rane, R.; Kulkarni, A.; Prajapati, H.; Taylor, R.; Jain, A.; Chen, V. Post-Process Effects of Isothermal Annealing and Initially Applied Static Uniaxial Loading on the Ultimate Tensile Strength of Fused Filament Fabrication Parts. *Materials* **2020**, *13* (2), 352.

(58) Koski, C.; Onuike, B.; Bandyopadhyay, A.; Bose, S. Starch-Hydroxyapatite Composite Bone Scaffold Fabrication Utilizing a Slurry Extrusion-Based Solid Freeform Fabricator. *Addit. Manuf.* **2018**, *24*, 47–59.

(59) Panwar, A.; Tan, L. P. Current Status of Bioinks for Micro-Extrusion-Based 3D Bioprinting. *Molecules* **2016**, *21* (6), 685.

(60) Qian, L.; Zhang, H. Controlled Freezing and Freeze Drying: A Versatile Route for Porous and Micro-/Nano-Structured Materials. *J. Chem. Technol. Biotechnol.* **2011**, *86* (2), 172–184. https://doi.org/10.1002/jctb.2495.

(61) Deville, S.; Saiz, E.; Nalla, R. K.; Tomsia, A. P. Freezing as a Path to Build Complex Composites. *Science* **2006**, *311* (5760), 515–518.

(62) Yang, X.-Y.; Chen, L.-H.; Li, Y.; Rooke, J. C.; Sanchez, C.; Su, B.-L. Hierarchically Porous Materials: Synthesis Strategies and Structure Design. *Chem. Soc. Rev.* **2017**, *46* (2), 481–558.

(63) Carlstedt, J.; Wojtasz, J.; Fyhr, P.; Kocherbitov, V. Understanding Starch Gelatinization: The Phase Diagram Approach. *Carbohydr. Polym.* **2015**, *129*, 62–69. https://doi.org/10.1016/j.carbpol.2015.04.045.

(64) Kowalski, S. J.; Musielak, G.; Rybicki, A.; Śliwa, T. Stresses and Strains in Elastic, Viscoelastic, and Plastic Materials during Drying. *Dry. Technol.* **2012**, 30 (11–12), 1176–1189. https://doi.org/10.1080/07373937.2012.692745.

(65) Dissa, A. O.; Desmorieux, H.; Savadogo, P. W.; Segda, B. G.; Koulidiati, J. Shrinkage, Porosity and Density Behaviour during Convective Drying of Spirulina. *J. Food Eng.* **2010**, *97* (3), 410–418. https://doi.org/10.1016/j.jfoodeng.2009.10.036.

(66) Umashankar, K.; Chandralekha, A.; Dandavate, T.; Tavanandi, H. A.; Raghavarao, K. A Nonconventional Method for Drying of Pseudomonas Aeruginosa and Its Comparison with Conventional Methods. *Dry. Technol.* **2019**, *37* (7), 839–853.

(67) Mueller, T.; Kusne, A. G.; Ramprasad, R. Machine Learning in Materials Science: Recent Progress and Emerging Applications. In *Reviews in Computational Chemistry, Volume 29*; Parrill, A. L., Lipkowitz, K. B., Series Eds.; Reviews in Computational Chemistry; John Wiley & Sons, Inc., 2016; Vol. 29, p 480.

(68) Attia, P. M.; Grover, A.; Jin, N.; Severson, K. A.; Markov, T. M.; Liao, Y.-H.; Chen, M. H.; Cheong, B.; Perkins, N.; Yang, Z.; Herring, P. K.; Aykol, M.; Harris, S. J.; Braatz, R. D.; Ermon, S.; Chueh, W. C. Closed-Loop Optimization of Fast-Charging Protocols for Batteries with

Machine Learning. *Nature* **2020**, *578* (7795), 397–402. https://doi.org/10.1038/s41586-020-1994-5.

(69) Yang, K.; Xu, X.; Yang, B.; Cook, B.; Ramos, H.; Krishnan, N. M. A.; Smedskjaer, M. M.; Hoover, C.; Bauchy, M. Predicting the Young's Modulus of Silicate Glasses Using High-Throughput Molecular Dynamics Simulations and Machine Learning. *Sci. Rep.* **2019**, *9* (1), 8739. https://doi.org/10.1038/s41598-019-45344-3.

(70) Xue, D.; Balachandran, P. V.; Hogden, J.; Theiler, J.; Xue, D.; Lookman, T. Accelerated Search for Materials with Targeted Properties by Adaptive Design. *Nat. Commun.* **2016**, *7* (1), 11241. https://doi.org/10.1038/ncomms11241.

(71) Zhang, Y.; Ling, C. A Strategy to Apply Machine Learning to Small Datasets in Materials Science. *Npj Comput. Mater.* **2018**, *4* (1), 25. https://doi.org/10.1038/s41524-018-0081-z.

(72) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559* (7715), 547–555. https://doi.org/10.1038/s41586-018-0337-2.

(73) Burkov, A. *Andriy Burkov's The Hundred-Page Machine Learning Book*; 2019.

(74) Selekman, J. A.; Qiu, J.; Tran, K.; Stevens, J.; Rosso, V.; Simmons, E.; Xiao, Y.; Janey, J. High-Throughput Automation in Chemical Process Development. *Annu. Rev. Chem. Biomol. Eng.* **2017**, *8* (1), 525–547. https://doi.org/10.1146/annurev-chembioeng-060816-101411.

(75) Mennen, S. M.; Alhambra, C.; Allen, C. L.; Barberis, M.; Berritt, S.; Brandt, T. A.; Campbell, A. D.; Castañón, J.; Cherney, A. H.; Christensen, M.; Damon, D. B.; Eugenio de Diego, J.; García-Cerrada, S.; García-Losada, P.; Haro, R.; Janey, J.; Leitch, D. C.; Li, L.; Liu, F.; Lobben, P. C.; MacMillan, D. W. C.; Magano, J.; McInturff, E.; Monfette, S.; Post, R. J.; Schultz, D.; Sitter, B. J.; Stevens, J. M.; Strambeanu, I. I.; Twilton, J.; Wang, K.; Zajac, M. A. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Org. Process Res. Dev.* **2019**, *23* (6), 1213–1242. https://doi.org/10.1021/acs.oprd.9b00140.

(76) Rodriguez, J.; Politi, M.; Scheiwiller, S.; Bonageri, S.; Adler, S.; Beck, D.; Pozzo, L. D. PhasIR: An Instrumentation and Analysis Software for High-Throughput Phase Transition Temperature Measurements. *J. Open Hardw.* **2021**, *5* (1). https://doi.org/10.5334/joh.39.

(77) Rodriguez Jr, J.; Politi, M.; Adler, S.; Beck, D.; Pozzo, L. High-Throughput and Data Driven Strategies for the Design of Deep-Eutectic Solvent Electrolytes. 16.

(78) Montgomery, D. C.; Runger, G. C. *Applied Statistics and Probability for Engineers*, 7th ed.; John Wiley & Sons, Inc., 2018.

(79) Box, G. E. P.; Hunter, J. S.; Hunter, W. G. *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd ed.; Wiley Series in Probability and Statistics; John Wiley & Sons, Inc., 2005.

(80) Li, C.; Rubín de Celis Leal, D.; Rana, S.; Gupta, S.; Sutti, A.; Greenhill, S.; Slezak, T.; Height, M.; Venkatesh, S. Rapid Bayesian Optimisation for Synthesis of Short Polymer Fiber Materials. *Sci. Rep.* **2017**, *7* (1), 5683. https://doi.org/10.1038/s41598-017-05723-0.

(81) Kerrigan, D.; Hullman, J.; Bertini, E. A Survey of Domain Knowledge Elicitation in Applied Machine Learning. *Multimodal Technol. Interact.* **2021**, *5* (12), 73. https://doi.org/10.3390/mti5120073.

(82) Deng, C.; Ji, X.; Rainey, C.; Zhang, J.; Lu, W. Integrating Machine Learning with Human Knowledge. *iScience* **2020**, *23* (11), 101656. https://doi.org/10.1016/j.isci.2020.101656.

(83) Häse, F.; Aldeghi, M.; Hickman, R. J.; Roch, L. M.; Aspuru-Guzik, A. Gryffin: An Algorithm for Bayesian Optimization of Categorical Variables Informed by Expert Knowledge. *Appl. Phys. Rev.* **2021**, *8* (3), 031406. https://doi.org/10.1063/5.0048164.

(84) Zhao, Q.; Yang, H.; Liu, J.; Zhou, H.; Wang, H.; Yang, W. Machine Learning-Assisted Discovery of Strong and Conductive Cu Alloys: Data Mining from Discarded Experiments and Physical Features. *Mater. Des.* **2021**, *197*, 109248. https://doi.org/10.1016/j.matdes.2020.109248.

(85) Ge, X.; Goodwin, R. T.; Yu, H.; Romero, P.; Abdelrahman, O.; Sudhalkar, A.; Kusuma, J.; Cialdella, R.; Garg, N.; Varshney, L. R. Accelerated Design and Deployment of Low-Carbon Concrete for Data Centers. *ArXiv220405397 Cs* **2022**.

(86) Severson, K. A.; Pfeiffer, O. P.; Chen, J.; Gong, K.; Gregory, J.; Goodwin, R.; Olivetti, E. A. Amortized Inference of Gaussian Process Hyperparameters for Improved Concrete Strength Trajectory Prediction. 10.

(87) Chen, C.; Gu, G. X. Generative Deep Neural Networks for Inverse Materials Design Using Backpropagation and Active Learning. *Adv. Sci.* **2020**, *7* (5), 1902607. https://doi.org/10.1002/advs.201902607.

(88) Gu, G. X.; Chen, C.-T.; Buehler, M. J. De Novo Composite Design Based on Machine Learning Algorithm. *Extreme Mech. Lett.* **2018**, *18*, 19–28. https://doi.org/10.1016/j.eml.2017.10.001.

(89) Gu, G. X.; Chen, C.-T.; Richmond, D. J.; Buehler, M. J. Bioinspired Hierarchical Composite Design Using Machine Learning: Simulation, Additive Manufacturing, and Experiment. *Mater. Horiz.* **2018**, *5* (5), 939–945. https://doi.org/10.1039/C8MH00653A.

(90) Yang, C.; Kim, Y.; Ryu, S.; Gu, G. X. Prediction of Composite Microstructure Stress-Strain Curves Using Convolutional Neural Networks. *Mater. Des.* **2020**, *189*, 108509. https://doi.org/10.1016/j.matdes.2020.108509.

(91) Wang, Y.; Zhang, Y.; Zhao, H.; Li, X.; Huang, Y.; Schadler, L. S.; Chen, W.; Brinson, L. C. Identifying Interphase Properties in Polymer Nanocomposites Using Adaptive Optimization. *Compos. Sci. Technol.* **2018**, *162*, 146–155. https://doi.org/10.1016/j.compscitech.2018.04.017.

(92) Diamantopoulou, M.; Karathanasopoulos, N.; Mohr, D. Stress-Strain Response of Polymers Made through Two-Photon Lithography: Micro-Scale Experiments and Neural Network Modeling. *Addit. Manuf.* **2021**, *47*, 102266. https://doi.org/10.1016/j.addma.2021.102266.

(93) Sakurai, A.; Yada, K.; Simomura, T.; Ju, S.; Kashiwagi, M.; Okada, H.; Nagao, T.; Tsuda, K.; Shiomi, J. Ultranarrow-Band Wavelength-Selective Thermal Emission with Aperiodic Multilayered Metamaterials Designed by Bayesian Optimization. *ACS Cent. Sci.* **2019**, *5* (2), 319–326. https://doi.org/10.1021/acscentsci.8b00802.

(94) Ma, W.; Cheng, F.; Xu, Y.; Wen, Q.; Liu, Y. Probabilistic Representation and Inverse Design of Metamaterials Based on a Deep Generative Model with Semi-Supervised Learning Strategy. *Adv. Mater.* **2019**, *31* (35), 1901111. https://doi.org/10.1002/adma.201901111.

(95) Erps, T.; Foshey, M.; Luković, M. K.; Shou, W.; Goetzke, H. H.; Dietsch, H.; Stoll, K.; von Vacano, B.; Matusik, W. Accelerated Discovery of 3D Printing Materials Using Data-Driven Multiobjective Optimization. *Sci. Adv.* **2021**, *7* (42), eabf7435. https://doi.org/10.1126/sciadv.abf7435.

(96) Park, S.; Baker, J. O.; Himmel, M. E.; Parilla, P. A.; Johnson, D. K. Cellulose Crystallinity Index: Measurement Techniques and Their Impact on Interpreting Cellulase Performance. *Biotechnol. Biofuels* **2010**, *3* (1), 1–10. https://doi.org/10.1186/1754-6834-3-10/TABLES/2.

(97) Dou, C.; Marcondes, W. F.; Djaja, J. E.; Bura, R.; Gustafson, R. Can We Use Short Rotation Coppice Poplar for Sugar Based Biorefinery Feedstock? Bioconversion of 2-Year-Old Poplar Grown as Short Rotation Coppice. *Biotechnol. Biofuels* **2017**, *10* (1), 144–144.

(98) Dou, C.; Gustafson, R.; Bura, R. Bridging the Gap between Feedstock Growers and Users: The Study of a Coppice Poplar-Based Biorefinery. *Biotechnol. Biofuels* **2018**, *11* (1), 77. https://doi.org/10.1186/s13068-018-1079-y.

(99) Hörhammer, H.; Dou, C.; Gustafson, R.; Suko, A.; Bura, R. Removal of Non-Structural Components from Poplar Whole-Tree Chips to Enhance Hydrolysis and Fermentation Performance. *Biotechnol. Biofuels* **2018**, *11* (1), 222. https://doi.org/10.1186/s13068-018-1219-4.

(100) VanLandingham, M. R.; Villarrubia, J. S.; Guthrie, W. F.; Meyers, G. F. Nanoindentation of Polymers: An Overview. *Macromol. Symp.* **2001**, *167* (1), 15–44. https://doi.org/10.1002/1521-3900(200103)167:1<15::AID-MASY15>3.0.CO;2-T.

(101) Campbell, A. C.; Buršíková, V.; Martinek, J.; Klapetek, P. Modeling the Influence of Roughness on Nanoindentation Data Using Finite Element Analysis. *Int. J. Mech. Sci.* **2019**, *161*, 105015.

(102) Böhme, L.; Keksel, A.; Ströer, F.; Bohley, M.; Kieren-Ehses, S.; Kirsch, B.; Aurich, J.; Seewig, J.; Kerscher, E. Micro Hardness Determination on a Rough Surface by Using Combined Indentation and Topography Measurements. *Surf. Topogr. Metrol. Prop.* **2019**, *7* (4), 045021.

(103) Walter, C.; Antretter, T.; Daniel, R.; Mitterer, C. Finite Element Simulation of the Effect of Surface Roughness on Nanoindentation of Thin Films with Spherical Indenters. *Surf. Coat. Technol.* **2007**, *202* (4–7), 1103–1107.

(104) Harris, C. R.; Millman, K. J.; Walt, S. J. van der; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; Kerkwijk, M. H. van; Brett, M.; Haldane, A.; Río, J. F. del; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array Programming with NumPy. *Nature* **2020**, *585* (7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2.

(105) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9* (3), 90–95. https://doi.org/10.1109/MCSE.2007.55.

(106) McKinney, W.; others. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*; Austin, TX, 2010; Vol. 445, pp 51–56.

(107) SciPy 1.0 Contributors; Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, İ.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17* (3), 261–272. https://doi.org/10.1038/s41592-019-0686-2.

(108) Bernaerts, T. M. M.; Gheysen, L.; Kyomugasho, C.; Jamsazzadeh Kermani, Z.; Vandionant, S.; Foubert, I.; Hendrickx, M. E.; Van Loey, A. M. Comparison of Microalgal Biomasses as Functional Food Ingredients: Focus on the Composition of Cell Wall Related Polysaccharides. *Algal Res.* **2018**, *32*, 150–161. https://doi.org/10.1016/j.algal.2018.03.017.

(109) Van Eykelenburg, C. On the Morphology and Ultrastructure of the Cell Wall of Spirulina Platensis. *Antonie Van Leeuwenhoek* **1977**, *43* (2), 89–99. https://doi.org/10.1007/BF00395664.

(110) Chen, W.; Xu, J.; Yu, Q.; Yuan, Z.; Kong, X.; Sun, Y.; Wang, Z.; Zhuang, X.; Zhang, Y.; Guo, Y. Structural Insights Reveal the Effective Spirulina Platensis Cell Wall Dissociation Methods for Multi-Output Recovery. *Bioresour. Technol.* **2020**, *300*, 122628. https://doi.org/10.1016/j.biortech.2019.122628.

(111) Phélippé, M.; Gonçalves, O.; Thouand, G.; Cogne, G.; Laroche, C. Characterization of the Polysaccharides Chemical Diversity of the Cyanobacteria Arthrospira Platensis. *Algal Res.* **2019**, *38*, 101426. https://doi.org/10.1016/j.algal.2019.101426.

(112) Sekharam, K. M.; Venkataraman, L. V.; Salimath, P. V. Structural Studies of a Glucan Isolated from Blue-Green Alga Spirulina Platensis. *Food Chem.* **1989**, *31* (2), 85–91. https://doi.org/10.1016/0308-8146(89)90019-8.

(113) Lee, R. E. *Phycology*, 5th ed.; Cambridge University Press, 2018. https://doi.org/10.1017/9781316407219.

(114) Djearamane, S.; Lim, Y. M.; Wong, L. S.; Lee, P. F. Cytotoxic Effects of Zinc Oxide Nanoparticles on Cyanobacterium Spirulina (Arthrospira) Platensis. *PeerJ* **2018**, *6*, e4682. https://doi.org/10.7717/peerj.4682.

(115) Larrosa, A. P. Q.; Comitre, A. A.; Vaz, L. B.; Pinto, L. A. A. Influence of Air Temperature on Physical Characteristics and Bioactive Compounds in Vacuum Drying of *Arthrospira Spirulina*: *Spirulina* Dried in Vacuum Dryer. *J. Food Process Eng.* **2017**, *40* (2), e12359. https://doi.org/10.1111/jfpe.12359.

(116) A Rapid and Non-Destructive Method for Quantifying Biomolecules in Spirulina Platensis via Fourier Transform Infrared – Attenuated Total Reflectance Spectroscopy. *Algal Res.* **2018**, *32*, 341–352. https://doi.org/10.1016/j.algal.2018.04.023.

(117) Sayadi, M. H.; Salmani, N.; Heidari, A.; Rezaei, M. R. Bio-Synthesis of Palladium Nanoparticle Using Spirulina Platensis Alga Extract and Its Application as Adsorbent. *Surf. Interfaces* **2018**, *10*, 136–143. https://doi.org/10.1016/j.surfin.2018.01.002.

(118) Larrosa, A. P. Q.; Camara, Á. S.; Moura, J. M.; Pinto, L. A. A. Spirulina Sp. Biomass Dried/Disrupted by Different Methods and Their Application in Biofilms Production. *Food Sci. Biotechnol.* **2018**, *27* (6), 1659–1665. https://doi.org/10.1007/s10068-018-0397-y.

(119) Saka, C.; Kaya, M.; Bekiroğullari, M. Spirulina Microalgal Strain as Efficient a Metal-free Catalyst to Generate Hydrogen via Methanolysis of Sodium Borohydride. *Int. J. Energy Res.* **2020**, *44* (1), 402–410. https://doi.org/10.1002/er.4936.

(120) Çelekli, A.; Yavuzatmaca, M.; Bozkurt, H. An Eco-Friendly Process: Predictive Modelling of Copper Adsorption from Aqueous Solution on Spirulina Platensis. *J. Hazard. Mater.* **2010**, *173* (1), 123–129. https://doi.org/10.1016/j.jhazmat.2009.08.057.

(121) Palanisamy, M.; Töpfl, S.; Berger, R. G.; Hertel, C. Physico-Chemical and Nutritional Properties of Meat Analogues Based on Spirulina/Lupin Protein Mixtures. *Eur. Food Res. Technol.* **2019**, *245* (9), 1889–1898. https://doi.org/10.1007/s00217-019-03298-w.

(122) Huber, T.; Najaf Zadeh, H.; Feast, S.; Roughan, T.; Fee, C. 3D Printing of Gelled and Cross-Linked Cellulose Solutions; an Exploration of Printing Parameters and Gel Behaviour. *Bioengineering* **2020**, *7* (2), 30.

(123) Hu, X.; Yang, Z.; Kang, S.; Jiang, M.; Zhou, Z.; Gou, J.; Hui, D.; He, J. Cellulose Hydrogel Skeleton by Extrusion 3D Printing of Solution. *Nanotechnol. Rev.* **2020**, *9* (1), 345–353.

(124) Cheng, Y.; Shi, X.; Jiang, X.; Wang, X.; Qin, H. Printability of a Cellulose Derivative for Extrusion-Based 3D Printing: The Application on a Biodegradable Support Material. *Front. Mater.* **2020**, *7*, 86.

(125) Miranda, J. A. T. de; Carvalho, L. M. J. de; Vieira, A. C. de M.; Castro, I. M. de. Scanning Electron Microscopy and Crystallinity of Starches Granules from Cowpea, Black and Carioca Beans in Raw and Cooked Forms. *Food Sci. Technol.* **2019**, *39*, 718–724.

(126) Meng, Q.; Li, B.; Li, T.; Feng, X.-Q. A Multiscale Crack-Bridging Model of Cellulose Nanopaper. *J. Mech. Phys. Solids* **2017**, *103*, 22–39. https://doi.org/10.1016/j.jmps.2017.03.004.

(127) Favier, V.; Canova, G.; Cavaillé, J.; Chanzy, H.; Dufresne, A.; Gauthier, C. Nanocomposite Materials from Latex and Cellulose Whiskers. *Polym. Adv. Technol.* **1995**, *6* (5), 351–355.

(128) Abeykoon, C.; Sri-Amphorn, P.; Fernando, A. Optimization of Fused Deposition Modeling Parameters for Improved PLA and ABS 3D Printed Structures. *Int. J. Lightweight Mater. Manuf.* **2020**, *3* (3), 284–297. https://doi.org/10.1016/j.ijlmm.2020.03.003.

(129) Tymrak, B. M.; Kreiger, M.; Pearce, J. M. Mechanical Properties of Components Fabricated with Open-Source 3-D Printers under Realistic Environmental Conditions. *Mater. Des.* **2014**, *58*, 242–246. https://doi.org/10.1016/j.matdes.2014.02.038.

(130) Domínguez-Rodríguez, G.; Ku-Herrera, J.; Hernández-Pérez, A. An Assessment of the Effect of Printing Orientation, Density, and Filler Pattern on the Compressive Performance of 3D Printed ABS Structures by Fuse Deposition. *Int. J. Adv. Manuf. Technol.* **2018**, *95* (5), 1685–1695.

(131) Wang, X.; Pang, Z.; Chen, C.; Xia, Q.; Zhou, Y.; Jing, S.; Wang, R.; Ray, U.; Gan, W.; Li, C.; Chen, G.; Foster, B.; Li, T.; Hu, L. All-Natural, Degradable, Rolled-Up Straws Based on Cellulose Micro- and Nano-Hybrid Fibers. *Adv. Funct. Mater.* **2020**, *30* (22), 1910417. https://doi.org/10.1002/adfm.201910417.

(132) Law, E. Bacterial Cellulose Nanocomposites with Lignin: Fabrication and Characterization of Structure, Mechanical & Thermal Properties, University of Washington, 2021.

(133) Kohli, R. Assessing Interaction Effects in Latin Square-Type Designs. *Int. J. Res. Mark.* **1988**, *5* (1), 25–37. https://doi.org/10.1016/0167-8116(88)90014-6.

(134) Wohlert, M.; Benselfelt, T.; Wågberg, L.; Furó, I.; Berglund, L. A.; Wohlert, J. Cellulose and the Role of Hydrogen Bonds: Not in Charge of Everything. *Cellulose* **2021**, 1–23.

(135) Benítez, A. J.; Torres-Rendon, J.; Poutanen, M.; Walther, A. Humidity and Multiscale Structure Govern Mechanical Properties and Deformation Modes in Films of Native Cellulose Nanofibrils. *Biomacromolecules* **2013**, *14* (12), 4497–4506. https://doi.org/10.1021/bm401451m.

(136) Aro, T.; Fatehi, P. Production and Application of Lignosulfonates and Sulfonated Lignin. *ChemSusChem* **2017**, *10* (9), 1861–1877. https://doi.org/10.1002/cssc.201700082.

(137) Farris, S.; Introzzi, L.; Biagioni, P.; Holz, T.; Schiraldi, A.; Piergiovanni, L. Wetting of Biopolymer Coatings: Contact Angle Kinetics and Image Analysis Investigation. *Langmuir* **2011**, *27* (12), 7563–7574. https://doi.org/10.1021/LA2017006/SUPPL_FILE/LA2017006_SI_001.PDF.

(138) Gomez, C.; Guardia, A.; Mantari, J. L.; Coronado, A. M.; Reddy, J. N. A Contemporary Approach to the MSE Paradigm Powered by Artificial Intelligence from a Review Focused on Polymer Matrix Composites. *Mech. Adv. Mater. Struct.* **2021**, 1–21. https://doi.org/10.1080/15376494.2021.1886379.

(139) Chen, C.-T.; Gu, G. X. Machine Learning for Composite Materials. *MRS Commun.* **2019**, *9* (2), 556–566. https://doi.org/10.1557/mrc.2019.32.

(140) Murphy, K. P. *Probabilistic Machine Learning: An Introduction*; Adaptive computation and machine learning series; The MIT Press: Cambridge, Massachusetts, 2022.

(141) Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian Reaction Optimization as a Tool for Chemical Synthesis. *Nature* **2021**, *590* (7844), 89–96. https://doi.org/10.1038/s41586-021-03213-y.

(142) Campbell, I. R.; Lin, M.-Y.; Iyer, H.; Parker, M.; Fredricks, J. L.; Liao, K.; Jimenez, A. M.; Grandgeorge, P.; Roumeli, E. Progress in Sustainable Polymers from Biological Matter. *Annu. Rev. Mater. Res.* **2023**, *53* (1), null. https://doi.org/10.1146/annurev-matsci-080921-083655.

(143) Karniadakis, G. E.; Kevrekidis, I. G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-Informed Machine Learning. *Nat. Rev. Phys.* **2021**, *3* (6), 422–440. https://doi.org/10.1038/s42254-021-00314-5.

(144) van Ee, A. F.; Poursartip, A.; Zobeiry, N. Theory-Guided Machine Learning Composites Processing Modelling for Manufacturability Assessment in Preliminary Design. **2019**.

APPENDIX A: MACHINE LEARNING CODE USING PYTHON

Import Statements

```python
import numpy as np
import pandas as pd
from scipy.stats import norm
import scipy.optimize
from sklearn.utils.optimize import _check_optimize_result


import matplotlib
import matplotlib.pyplot as plt
from matplotlib import cm
from mpl_toolkits.mplot3d import Axes3D
from mpl_toolkits.mplot3d.art3d import Poly3DCollection
from mplcursors import cursor
matplotlib.use('Qt5Agg')

import torch
import gpytorch
import sys
import math

from sklearn.model_selection import train_test_split
from sklearn.gaussian_process import GaussianProcessRegressor
from sklearn.gaussian_process.kernels import RBF
from sklearn.model_selection import GridSearchCV

#unnecessary plotting libraries to figure out what works best.
import plotly.figure_factory as ff
import ternary
import plotly.graph_objects as go
```

Reading in Data

```python
import random

data = pd.read_csv("D:\Research\ThirdProject_ML\data_sets\ML_feature_data.csv")
groups = data.groupby(["chlorella mass (wt. proportion)", "dry BC mass (wt. proportion)", "SA mass (wt. proportion)"]).mean()

feature_groups = np.asarray([np.asarray(i) for i in groups.index.values]).squeeze()

inputs = data[["chlorella mass (wt. proportion)", "dry BC mass (wt. proportion)", "SA mass (wt. proportion)"]].values
strength = data[["Strength_MPa"]].values
EtB = data[["EtB_percent (max)"]].values
toughness = data[["Toughness_MJ_m3 (max)"]].values

random.seed(8)

test_size=0.1

features_for_train = random.sample(feature_groups.tolist(), k=len(feature_groups)- math.ceil(test_size*len(feature_groups)))

idx_train=[]
for i in range(0,len(features_for_train)):
    idx_temp = np.where((inputs == features_for_train[i]).all(1))[0]
    idx_train = np.append(idx_train,idx_temp).astype(int)

strength_train = np.array(strength[idx_train])
toughness_train = np.array(toughness[idx_train])
EtB_train = np.array(EtB[idx_train])
inputs_train = np.array(inputs[idx_train])
inputs_test = np.array(np.delete(inputs, idx_train, axis=0))
strength_test = np.array(np.delete(strength, idx_train, axis=0))
toughness_test = np.array(np.delete(toughness, idx_train, axis=0))
EtB_test = np.array(np.delete(EtB, idx_train, axis=0))

mean_strength_train = np.nanmean(strength_train)
std_strength_train = np.nanstd(strength_train)
mean_toughness_train = np.nanmean(toughness_train)
std_toughness_train = np.nanstd(toughness_train)
mean_EtB_train = np.nanmean(EtB_train)
std_EtB_train = np.nanstd(EtB_train)

mean_inputs_train = np.nanmean(inputs_train, axis=0)
```

Reading Data and Exact GP Model Training

```python
#import the data generated earlier (possible to change this to another file
in the future)
data_dict = torch.load('chlorellaSABC_preprocessed.pt')

#Unpack the data
mean_strength_train = data_dict['mean_strength_train']
std_strength_train = data_dict['std_strength_train']
mean_toughness_train = data_dict['mean_toughness_train']
std_toughness_train = data_dict['std_toughness_train']
mean_EtB_train = data_dict['mean_EtB_train']
std_EtB_train = data_dict['std_EtB_train']

mean_inputs_train = data_dict['mean_inputs_train']
std_inputs_train = data_dict['std_inputs_train']

train_inputs = data_dict['inputs_train_']
train_strength = data_dict['strength_train_'].squeeze()
train_toughness = data_dict['toughness_train_'].squeeze()
train_EtB = data_dict['EtB_train_'].squeeze()

test_inputs = data_dict['inputs_test_']
test_strength = data_dict['strength_test_'].squeeze()
test_toughness = data_dict['toughness_test_'].squeeze()
test_EtB = data_dict['EtB_test_'].squeeze()

# We will use the simplest form of GP model, exact inference
class ExactGPModel(gpytorch.models.ExactGP):
    def __init__(self, train_inputs, train_property, likelihood):
        super(ExactGPModel, self).__init__(train_inputs, train_property,
likelihood)
        self.mean_module = gpytorch.means.ConstantMean()
        self.covar_module =
gpytorch.kernels.ScaleKernel(gpytorch.kernels.RBFKernel())

    def forward(self, x):
        mean_x = self.mean_module(x)
        covar_x = self.covar_module(x)
        return gpytorch.distributions.MultivariateNormal(mean_x, covar_x)

# initialize likelihood and model
likelihood = gpytorch.likelihoods.GaussianLikelihood()
# model = ExactGPModel(train_inputs, train_strength, likelihood)
# model = ExactGPModel(train_inputs, train_toughness, likelihood)
model = ExactGPModel(train_inputs, train_EtB, likelihood)

```

Definitions of acquisition and triangle mesh functions

```python
# Define an acquisition function using the Expected Improvement algorithm
def custom_acquisition_function(X, model, best_value, xi=0.01):
    # Reshape the input compositions
    X_reshaped = np.column_stack([X[0].ravel(), X[1].ravel(), X[2].ravel()])

    # Calculate the predicted mean and standard deviation using the model
    predicted_mean, predicted_std = gpr.predict(X_reshaped, return_std=True)

    # Calculate improvement over the current best value
    improvement = predicted_mean - best_value

    # Calculate Z-score and expected improvement
    z = improvement / (predicted_std + 1e-9)  # Adding small constant for numerical stability
    ei = improvement * norm.cdf(z) + predicted_std * norm.pdf(z)

    # Add exploration parameter
    ei_reshaped = ei.reshape(X[0].shape)

    return ei_reshaped
```

```python
# Usage example:
# X: Candidate points
# model: Gaussian process model
# best_value: Current best observed value

#for optimizing a product of two properties - not used for Jeremy's dissertation work
def custom_acquisition_function2(X, model1, model2, best_value, xi=0.01):
    # Reshape the input compositions
    X_reshaped = np.column_stack([X[0].ravel(), X[1].ravel(), X[2].ravel()])

    # Calculate the predicted mean for each property using the respective models
    predicted_mean1 = model1.predict(X_reshaped)
    predicted_mean2 = model2.predict(X_reshaped)

    # Calculate the product of the properties as the objective function
    objective = predicted_mean1 * predicted_mean2

    # Calculate improvement over the current best value
    improvement = objective - best_value

    # Add exploration parameter
    ei_reshaped = improvement.reshape(X[0].shape)

    return ei_reshaped
```

```python
# Define the vertices for the triangles
def trianglemesh(c1_mesh, c2_mesh, point_mesh):
    triangles = []
    n_points = 50
    for i in range(n_points - 1):
        for j in range(n_points - 1):
            vertices = [
                (c1_mesh[i, j], c2_mesh[i, j], point_mesh[i, j]),
                (c1_mesh[i, j + 1], c2_mesh[i, j + 1], point_mesh[i, j + 1]),
                (c1_mesh[i + 1, j], c2_mesh[i + 1, j], point_mesh[i + 1, j])
            ]
            triangles.append(vertices)
            vertices = [
                (c1_mesh[i, j + 1], c2_mesh[i, j + 1], point_mesh[i, j + 1]),
                (c1_mesh[i + 1, j + 1], c2_mesh[i + 1, j + 1], point_mesh[i + 1, j + 1]),
                (c1_mesh[i + 1, j], c2_mesh[i + 1, j], point_mesh[i + 1, j])
            ]
            triangles.append(vertices)

    return triangles
```

Model for elongation at max stress (EtB)

```
#EtB curves - cleaned code for publication
# Generate some example training data
train_compositions = train_inputs
train_properties = train_EtB

# Define the Gaussian Process Regression model
kernel = RBF(length_scale=0.299, length_scale_bounds = "fixed")
gpr = GaussianProcessRegressor(kernel=kernel, optimizer='fmin_l_bfgs_b')

# Train the model
gpr.fit(unscore(train_compositions,mean_inputs_train, std_inputs_train), unscore(train_properties,mean_EtB_train, std_EtB_train)) #unscored values

# Generate a meshgrid of compositions for plotting
n_points = 50
c1 = np.linspace(0, 1, n_points)
c2 = np.linspace(0, 1, n_points)
c1_mesh, c2_mesh = np.meshgrid(c1, c2)
print("c1,2 mesh shape:", c1_mesh.shape, c2_mesh.shape)
positivevaluemask = c1_mesh + c2_mesh <= 1
c1_meshmasked, c2_meshmasked = c1_mesh[positivevaluemask], c2_mesh[positivevaluemask]
print("c1,2 masked mesh shape:", c1_meshmasked.shape, c2_meshmasked.shape)
c3_mesh = 1 - c1_mesh - c2_mesh
print("c3 mesh shape:", c3_mesh.shape)
c3_meshmasked = 1 - c1_meshmasked - c2_meshmasked
print("c3 masked mesh shape:", c3_meshmasked.shape)

# Flatten the meshgrid into a 2D array of compositions
test_compositions1 = np.column_stack([c1_mesh.ravel(), c2_mesh.ravel(), c3_mesh.ravel()])
test_compositions2 = np.column_stack([c1_meshmasked.ravel(), c2_meshmasked.ravel(), c3_meshmasked.ravel()])
print("stacked mesh shapes:", test_compositions1.shape,test_compositions2.shape)
```

```
# Predict the output properties for the test compositions
predicted_properties1, predicted_std1 = gpr.predict(test_compositions1,
return_std=True)
predicted_properties2, predicted_std2 = gpr.predict(test_compositions2,
return_std=True)
print("predicted properties/std (1) shape:", predicted_properties1.shape,
predicted_std1.shape)
print("predicted properties/std (2) shape:", predicted_properties2.shape,
predicted_std2.shape)
print(f"Kernel parameters before fit: \n{kernel})")
print(f"Kernel parameters after fit: \n{gpr.kernel_} \n")
# Calculate upper and lower confidence bounds
confidence_level = 0.95
z_value = norm.ppf((1 + confidence_level) / 2)  # Z-value for desired
confidence level
lower_bound = predicted_properties1 - z_value * predicted_std1
upper_bound = predicted_properties1 + z_value * predicted_std1

# Reshape the predicted properties and standard deviations back to the
meshgrid shape
predicted_properties_mesh1 = predicted_properties1.reshape(c1_mesh.shape)
max_indices=np.argwhere(np.max(predicted_properties_mesh1)==predicted_propert
ies_mesh1)
print(np.column_stack((c1_mesh[max_indices[:, 0], max_indices[:, 1]],
                                  c2_mesh[max_indices[:, 0], max_indices[:,
1]])))

# Reshape the bounds arrays to match the meshgrid shape
lower_bound_mesh = lower_bound.reshape(c1_mesh.shape)
upper_bound_mesh = upper_bound.reshape(c1_mesh.shape)

# Calculate the acquisition function values
acq_values = custom_acquisition_function((c1_mesh,c2_mesh,c3_mesh), gpr,
np.max(predicted_properties_mesh1))

# Find the index of the highest acquisition function value
best_index = np.argmax(acq_values)

# Get the corresponding coordinates from the grid
best_coords = (c1_mesh.ravel()[best_index], c2_mesh.ravel()[best_index])

# Print the suggested coordinates
print("Next Best Coordinates:", best_coords)
print("Acquisition value:", np.max(acq_values))
```

```
print("Prediction at the test data composition: ", "(C=0.15, BC=0.65, SA =
0.20)", gpr.predict(np.array([.15, .65, .20]).reshape(1,-1))) #composition of
the test data

# Plot the results in a 3D triangular plot
fig1 = plt.figure()
ax1 = fig1.add_subplot(111, projection='3d')
poly_collection1 =
Poly3DCollection(trianglemesh(c1_mesh,c2_mesh,predicted_properties_mesh1),
alpha=0.2, facecolors='blue', picker = True)
ax1.add_collection3d(poly_collection1)
ax1.scatter(unscore(train_compositions[:, 0], mean_inputs_train[0],
std_inputs_train[0]).numpy(),
            unscore(train_compositions[:, 1],mean_inputs_train[1],
std_inputs_train[1]).numpy(),
            unscore(train_properties,mean_EtB_train, std_EtB_train).numpy(),
c='red', marker='o', label='Training Data')
ax1.scatter(unscore(test_inputs[:,0],mean_inputs_train[0],
std_inputs_train[0]).numpy(),
            unscore(test_inputs[:,1],mean_inputs_train[1],
std_inputs_train[1]).numpy(),
            unscore(test_EtB, mean_EtB_train, std_EtB_train).numpy(),
            'r*', label='Test Data')    # Plot predictive means as blue line

ax1.set_xlabel('Chlorella wt. proportion')
ax1.set_ylabel('dry BC wt. proportion')
ax1.set_zlabel('EtB')
ax1.legend()

# Set perspective and limits
ax1.view_init(elev=20, azim=-20)
ax1.set_xlim(0, 1)
ax1.set_ylim(0, 1)
ax1.invert_xaxis()

# Add cursor hover to display coordinates and properties
# cursor1 = cursor(ax1, hover=True)

# Add an acquisition function plot
# Reshape the acquisition function values to match the meshgrid shape
acq_values_mesh = acq_values.reshape(c1_mesh.shape)
fig2 = plt.figure()
ax2 = fig2.add_subplot(111, projection='3d', frame_on=False)
ax2.plot_surface(c1_mesh, c2_mesh, acq_values_mesh, cmap='cool', alpha=0.5)
ax2.set_zticks([])
```

```
# Set perspective and limits for the acquisition function plot
ax2.view_init(elev=20, azim=-40)
ax2.set_xlim(0, 1)
ax2.set_ylim(0, 1)

# Add cursor hover to display coordinates and properties
# cursor2 = cursor(ax2, hover=True)

fig7, ax7 = plt.subplots()

# Set the axis labels
ax7.set_xlabel('Chlorella wt. proportion')
ax7.set_ylabel('dry BC wt. proportion')

# Create the contour plot
contour = ax7.tricontourf(c1_mesh.flatten(), c2_mesh.flatten(),
predicted_properties_mesh1.flatten(), levels=10, cmap='viridis')

# Add a colorbar
cbar = fig7.colorbar(contour)
cbar.set_label('EtB')

# Show the plot
plt.show()

# Show the plots in separate interactive windows
plt.show(block=False)
plt.figure(fig1.number)  # Activate the first figure window
plt.show()
```

Model for toughness

```
#toughness
train_compositions = train_inputs
train_properties = train_toughness

# Define the Gaussian Process Regression model
kernel = RBF(length_scale=0.196, length_scale_bounds="fixed") #0.196 is ideal
given our data points
gpr = GaussianProcessRegressor(kernel=kernel, optimizer='fmin_l_bfgs_b')

# Train the model
gpr.fit(unscore(train_compositions,mean_inputs_train, std_inputs_train),
unscore(train_properties,mean_toughness_train, std_toughness_train))
#unscored values

# Generate a meshgrid of compositions for plotting
n_points = 50
c1 = np.linspace(0, 1, n_points)
c2 = np.linspace(0, 1, n_points)
c1_mesh, c2_mesh = np.meshgrid(c1, c2)
print("c1,2 mesh shape:", c1_mesh.shape, c2_mesh.shape)
c3_mesh = 1 - c1_mesh - c2_mesh
print("c3 mesh shape:", c3_mesh.shape)


# Flatten the meshgrid into a 2D array of compositions
test_compositions1 = np.column_stack([c1_mesh.ravel(), c2_mesh.ravel(),
c3_mesh.ravel()])

# Predict the output properties for the test compositions
predicted_properties1, predicted_std1 = gpr.predict(test_compositions1,
return_std=True)
print("predicted properties/std (1) shape:", predicted_properties1.shape,
predicted_std1.shape)
print(f"Kernel parameters before fit: \n{kernel})")
print(f"Kernel parameters after fit: \n{gpr.kernel_} \n")
# Calculate upper and lower confidence bounds
confidence_level = 0.95
z_value = norm.ppf((1 + confidence_level) / 2)  # Z-value for desired
confidence level
lower_bound = predicted_properties1 - z_value * predicted_std1
upper_bound = predicted_properties1 + z_value * predicted_std1
```

```
# Reshape the predicted properties and standard deviations back to the meshgrid shape
predicted_properties_mesh1 = predicted_properties1.reshape(c1_mesh.shape)
max_indices=np.argwhere(np.max(predicted_properties_mesh1)==predicted_properties_mesh1)
print(np.column_stack((c1_mesh[max_indices[:, 0], max_indices[:, 1]],
                                  c2_mesh[max_indices[:, 0], max_indices[:, 1]])))
# Reshape the bounds arrays to match the meshgrid shape
lower_bound_mesh = lower_bound.reshape(c1_mesh.shape)
upper_bound_mesh = upper_bound.reshape(c1_mesh.shape)

# Calculate the acquisition function values
acq_values = custom_acquisition_function((c1_mesh,c2_mesh,c3_mesh), gpr, np.max(predicted_properties_mesh1))

# Find the index of the highest acquisition function value
best_index = np.argmax(acq_values)

# Get the corresponding coordinates from the grid
best_coords = (c1_mesh.ravel()[best_index], c2_mesh.ravel()[best_index])

# Print the suggested coordinates
print("Next Best Coordinates:", best_coords)
print("Acquisition value:", np.max(acq_values))
print("Prediction at the test data composition: ", "(C=0.15, BC=0.65, SA = 0.20)", gpr.predict(np.array([.15, .65, .20]).reshape(1,-1)))

# Plot the results in a 3D triangular plot
fig1 = plt.figure()
ax1 = fig1.add_subplot(111, projection='3d')
poly_collection1 = Poly3DCollection(trianglemesh(c1_mesh,c2_mesh,predicted_properties_mesh1), alpha=0.2, facecolors='blue', picker = True)
ax1.add_collection3d(poly_collection1)
```

```
ax1.scatter(unscore(train_compositions[:, 0], mean_inputs_train[0],
std_inputs_train[0]).numpy(),
            unscore(train_compositions[:, 1],mean_inputs_train[1],
std_inputs_train[1]).numpy(),
            unscore(train_properties,mean_toughness_train,
std_toughness_train).numpy(), c='red', marker='o', label='Training Data')
ax1.scatter(unscore(test_inputs[:,0],mean_inputs_train[0],
std_inputs_train[0]).numpy(),
            unscore(test_inputs[:,1],mean_inputs_train[1],
std_inputs_train[1]).numpy(),
            unscore(test_toughness, mean_toughness_train,
std_toughness_train).numpy(),
            'r*', label='Test Data')    # Plot predictive means as blue line

ax1.set_xlabel('Chlorella wt. proportion')
ax1.set_ylabel('dry BC wt. proportion')
ax1.set_zlabel('toughness')
ax1.legend()

# Set perspective and limits
ax1.view_init(elev=20, azim=-20)
ax1.set_xlim(0, 1)
ax1.set_ylim(0, 1)
# ax1.set_zlim(0, np.max(predicted_properties_mesh))
ax1.invert_xaxis()

# Add cursor hover to display coordinates and properties
# cursor1 = cursor(ax1, hover=True)

# Add an acquisition function plot
# Reshape the acquisition function values to match the meshgrid shape
acq_values_mesh = acq_values.reshape(c1_mesh.shape)
fig2 = plt.figure()
ax2 = fig2.add_subplot(111, projection='3d', frame_on=False)
ax2.plot_surface(c1_mesh, c2_mesh, acq_values_mesh, cmap='cool', alpha=0.5)
ax2.set_zticks([])

# Set perspective and limits for the acquisition function plot
ax2.view_init(elev=20, azim=-40)
ax2.set_xlim(0, 1)
ax2.set_ylim(0, 1)
```

```
# Add cursor hover to display coordinates and properties
# cursor2 = cursor(ax2, hover=True)

fig7, ax7 = plt.subplots()

# Set the axis labels
ax7.set_xlabel('Chlorella wt. proportion')
ax7.set_ylabel('dry BC wt. proportion')

# Create the contour plot
contour = ax7.tricontourf(c1_mesh.flatten(), c2_mesh.flatten(),
predicted_properties_mesh1.flatten(), levels=10, cmap='viridis')

# Add a colorbar
cbar = fig7.colorbar(contour)
cbar.set_label('toughness (MJ/m^3)')

# Show the plot
plt.show()

# Show the plots in separate interactive windows
plt.show(block=False)
plt.figure(fig1.number)  # Activate the first figure window
plt.show()
```

Length scale vs predicted value plot

```
#length scale vs predicted value
length_scales = np.linspace(0.001, 1.000, num=999)  # Generate an array of
length_scale values
predictions = []

for length_scale in length_scales:
    # Create and fit the Gaussian Process Regressor with the current
length_scale
    kernel = RBF(length_scale=length_scale, length_scale_bounds="fixed")
    gpr = GaussianProcessRegressor(kernel=kernel, optimizer='fmin_l_bfgs_b')
    gpr.fit(unscore(train_compositions, mean_inputs_train, std_inputs_train),
            unscore(train_properties, mean_toughness_train,
std_toughness_train))

    # Predict at the specific composition
    predictions.append(gpr.predict(np.array([.15, 0.65, .2]).reshape(1,-1))-
1.701827176)

# Convert predictions to a numpy array for plotting
predictions = np.array(predictions)
# Find the index where gpr.predict is closest to zero
zero_index = np.argmin(np.abs(predictions))

# Plot predictions vs. length_scale
plt.plot(length_scales, predictions)
plt.scatter(length_scales[zero_index], predictions[zero_index], color='red',
label='Zero Prediction')
plt.xlabel('Length Scale')
plt.ylabel('Predicted Value Difference')
plt.title('Difference in Predicted Value with Length Scale')
plt.legend()

# Add text annotations for the coordinates
text = f'({length_scales[zero_index]:.3f}, {predictions[zero_index][0]:.3f})'
plt.text(length_scales[zero_index], predictions[zero_index], text,
         ha='center', va='bottom', color='red')

# Print the length_scale value where gpr.predict is closest to zero
print("Length Scale at Zero Prediction:", length_scales[zero_index])
plt.show()


# print(gpr.predict(np.array([.15, 0.65, .2]).reshape(1,-1))-1.701827176)
#this number is the mean toughness
```

Ternary Contour Plots

```
#TERNARY CONTOUR PLOTS

from plotly.offline import plot

a = test_compositions1[:, 0]
b = test_compositions1[:, 1]
mask = a + b <= 1
a, b = a[mask], b[mask]
c = 1 - a - b

fig3 = ff.create_ternary_contour(
    [a,b,c],
    predicted_properties1[mask],
    interp_mode='cartesian',
    ncontours=20,
    colorscale='Viridis',
    showscale = True,
    coloring='lines',
    #showmarkers = True
)

specific_points_train = [
    unscore(train_compositions[:, 0], mean_inputs_train[0],
std_inputs_train[0]).numpy(),
    unscore(train_compositions[:, 1], mean_inputs_train[1],
std_inputs_train[1]).numpy(),
    unscore(train_compositions[:, 2], mean_inputs_train[2],
std_inputs_train[2]).numpy(),
]

specific_points_test = [
    unscore(test_inputs[:,0], mean_inputs_train[0],
std_inputs_train[0]).numpy(),
    unscore(test_inputs[:,1],mean_inputs_train[1],
std_inputs_train[1]).numpy(),
    unscore(test_inputs[:,2],mean_inputs_train[2],
std_inputs_train[2]).numpy(),
]
```

```
scatter_trace_train = go.Scatterternary(
    a=specific_points_train[0],
    b=specific_points_train[1],
    c=specific_points_train[2],
    mode='markers',
    marker=dict(
        size=15,
        color='red',
        symbol='circle'
    ),
    name='Specific Points',
)

scatter_trace_test = go.Scatterternary(
    a=specific_points_test[0],
    b=specific_points_test[1],
    c=specific_points_test[2],
    mode='markers',
    marker=dict(
        size=8,
        color='blue',
        symbol='circle'
    ),
    name='Specific Points'
)

fig3.add_trace(scatter_trace_train)
fig3.add_trace(scatter_trace_test)
```

```
# Update the layout
fig3.update_layout(
    title='Ternary Contour Plot - Elongation to Break',
    # title='Ternary Contour Plot - Toughness',
    ternary=dict(sum=1,
                 aaxis=dict(
                     title='chlorella wt. proportion',
                     tickvals=[-.1, 0, 0.2, 0.4, 0.6, 0.8, 1, 1.1],  # Adjust the tick values for the a-axis
                     ticktext=[-.1, 0, 0.2, 0.4, 0.6, 0.8, 1, 1.1]  # Adjust the tick labels for the a-axis
                ),
                 baxis=dict(
                     title='dry BC wt. proportion',
                     tickvals=[-.1, 0, 0.2, 0.4, 0.6, 0.8, 1, 1.1],  # Adjust the tick values for the a-axis
                     ticktext=[-.1, 0, 0.2, 0.4, 0.6, 0.8, 1, 1.1]  # Adjust the tick labels for the a-axis
                ),
                 caxis=dict(
                     title='stearic acid wt. proportion',
                     tickvals=[-.1, 0, 0.2, 0.4, 0.6, 0.8, 1, 1.1],  # Adjust the tick values for the a-axis
                     ticktext=[-.1, 0, 0.2, 0.4, 0.6, 0.8, 1, 1.1]  # Adjust the tick labels for the a-axis
                )
    ),
    autosize=False,
    width=600,
    height=600,
)


plt.tight_layout()
# Show the ternary contour plot
fig3.show()
# plot(fig3, filename='ternary_contour_plot.html', auto_open=True)

# Find the maximum value and its index
max_value = np.max(predicted_properties1[mask])
max_index = np.argmax(predicted_properties1[mask])
```

```
# Get the corresponding composition values
max_a = a[max_index]
max_b = b[max_index]
max_c = c[max_index]

print("Maximum Value:", max_value)
print("Composition (a, b, c):", max_a, max_b, max_c)

from scipy.signal import argrelextrema
# Find all local maxima indices
maxima_indices = argrelextrema(predicted_properties1[mask], np.greater)

# Get the corresponding composition values and predicted property values of
local maxima
maxima_a = a[maxima_indices]
maxima_b = b[maxima_indices]
maxima_c = c[maxima_indices]
maxima_values = predicted_properties1[mask][maxima_indices]

# Sort the list of local maxima by max values in descending order
sorted_indices = np.argsort(maxima_values, axis=None)[::-1]
sorted_maxima_values = maxima_values[sorted_indices]
sorted_maxima_a = maxima_a[sorted_indices]
sorted_maxima_b = maxima_b[sorted_indices]
sorted_maxima_c = maxima_c[sorted_indices]
```

```
# Check if there are at least two local maxima
if len(sorted_maxima_values) >= 4:
    # Output the second highest value
    second_highest_value = sorted_maxima_values[1]
    third_highest_value = sorted_maxima_values[2]
    fourth_highest_value = sorted_maxima_values[3]

    second_highest_a = sorted_maxima_a[1]
    second_highest_b = sorted_maxima_b[1]
    second_highest_c = sorted_maxima_c[1]

    third_highest_a = sorted_maxima_a[2]
    third_highest_b = sorted_maxima_b[2]
    third_highest_c = sorted_maxima_c[2]

    fourth_highest_a = sorted_maxima_a[3]
    fourth_highest_b = sorted_maxima_b[3]
    fourth_highest_c = sorted_maxima_c[3]

    print("Second Highest Value:", second_highest_value)
    print("Composition (a, b, c):", second_highest_a, second_highest_b,
second_highest_c)
    print()

    print("Third Highest Value:", third_highest_value)
    print("Composition (a, b, c):", third_highest_a, third_highest_b,
third_highest_c)
    print()

    print("Fourth Highest Value:", fourth_highest_value)
    print("Composition (a, b, c):", fourth_highest_a, fourth_highest_b,
fourth_highest_c)
else:
    print("There are less than four local maxima.")
```