# Topological Analysis of Fuel Structure Data

**Name:** Olivia Chandrasekhar
**Hosting Site:** USDA Forest Service Southern Research Station
**Mentor:** Dr. Louise Loudermilk

**Mentor's Signature:** ________________________________

**FIGURE 1.** In the field with fellow NSF MSGI intern Alex Mendez. Picture: Dr. Louise Loudermilk


## Abstract

Detailed information about the three-dimensional structure of vegetative fuels is an indispensable component of computational models of fire behavior. In particular, fine-scale heterogeneity in fuels plays an important role in the behavior of prescribed fire. As such, the need exists for improved methodologies to model fine scale fuels and clarify the link between fuel structures at different scales. In this project, we apply topological data analysis to study the persistent homology of LiDAR scans taken in the Hitchiti Experimental Forest. We identify appropriate filtration levels for different kinds of data, evaluate several metrics of the connectivity of our data sets, and develop a means of classifying scans by relative connectivity. Physics-based models of fire behavior rely on accurate fuel inputs at appropriate scales. As such, this work will benefit the community of fire scientists employing multi-scale, multi-physics models for simulation and prediction.

# 1. INTRODUCTION

Fire behavior is intrinsically linked to fuel structure, as discussed in numerous studies, including [1]. In particular, the behavior of prescribed fire–which burns at a low intensity corresponding to lower flame height–is affected by changes in fuel structure at a very fine scale. As such, detailed, three-dimensional information about fine-scale fuel heterogeneity is a necessary component in accurate fire modeling and simulation. Unfortunately, collecting this data is a time and resource expensive task. As a result, there is a need for accurate modeling methodologies to populate fuel data in computational models of fire propagation.

# 2. DESCRIPTION OF PROJECT

2.1. **Background.** Both Los Alamos National Lab's fully physics-based CFD model FIRETEC and its light-weight, cellular automata-based counterpart QUIC-fire use four fuel characteristics as input: bulk density, water mass per volume, particle size and fuel height [2]. These metrics are based on an understanding of the spatial distribution of the various kinds of plant matter that constitute fuels in the region of interest. The need to more effectively characterize this vegetation, particularly in a way that captures its three-dimensional structure at a fine scale, is addressed in [3], [4], [2], and others.

For this project, we used data obtained from the Hitchiti Experimental Forest, a nearly $5,000$ acre US Forest Service Research site in Jones Country, Georgia. The Hitchiti is a Piedmont landscape consisting of upland Loblolly pine and Sweetgum and bottomland hardwoods. The Hitchiti is home to the endangered Red-cockaded woodpecker, whose habitat is dependent on frequent burning. Prescribed burns are conducted frequently, with most areas on a two- to five-year burning cycle. The Hitchiti is named after its original inhabitants, the Hitchiti Indians, who were also the first to conduct controlled burns in the region [5].

Vegetation data available for this region takes two forms: terrestrial and aerial LiDAR (Light Detection and Ranging) data. Terrestrial LiDAR scan (TLS) data is collected using $360°$ BLK scans with a $12$ meter radius taken on the forest floor, and the resulting scan has a a resolution of $10-15$ points per square centimeter, depending on the range of the scanner. This data is detailed and captures fine-scale ($< 1$ m) features of the vegetation, but is harder to collect at landscape scale. For more details on terrestrial scanning, see [1]. Aerial LiDAR scans (ALS) are conducted via airplane and cover the entire burn unit. Each ALS scan covers approximately $1,000$ square meters. At a resolution of $15-20$ points per square meter, the ALS does not capture the same level of detail as TLS, and is best used to characterize the structure of the canopy, as it may not penetrate to surface fuels. ALS scans are, however, readily available in many places. As a result, developing methods for populating surface fuels given ALS data is an important step in the effort to make fire modeling tools more accessible to land managers.

2.2. **Research Question.** ALS data captures large-scale features and is available at landscape scale, whereas TLS data captures fine-scale features in a much smaller area. The link between fuel characteristics observed at different scales is not well understood. This leads to the following research question: how do we best scale TLS data to landscape scale? Can we do this in a way that is informed by the ALS data? For this project, we used a topological approach to analyzing the data, which allowed us to turn this question about fuels in to a question about topology. In particular, we sought to find a correlation between topological features found in the ALS data and those found in the TLS data–in other words, topological features across scales. In practice, this is a three part question:

(1) Can we identify large scale patterns in the ALS data using topological analysis?
(2) Can we identify fine scale structure in the TLS data using topological analysis?
(3) Can we correlate the topological features we observe across scales, either with each other or with known ecological features?

2.3. **Topology Basics.** In order understand the terms used throughout the remainder of this report, we require some mathematical definitions.

**Def 2.1.** A topology (or topological space) is an ordered pair $(\chi, \tau)$, where $\chi$ is a set and $\tau$ is a set of open subsets of $\chi$. These subsets must satisfy three properties:

(1) The empty set and $\chi$ are subsets of $\tau$.
(2) The union of any set of elements of $\tau$ is also in $\tau$.
(3) The intersection of any finite set of elements in $\tau$ is in $\tau$.

**Def 2.2.** A metric space is an ordered pair $(V, d)$ where $V$ is a set and $d$ is a metric defined for elements of that set. A common example of a metric space is $(\mathbb{R}^2, d)$ where $d(x, y) = \sqrt{x^2 + y^2}$, the Euclidean norm.

**Def 2.3.** In the language of the previous two definitions, a metric topology is a topology on the space $V$ induced by the norm $d$, in which the open sets $\tau$ of definition 2.1 are open balls of radius $r$, where this radius is defined in terms of the metric $d$.

**Def 2.4.** Formally, we define these open balls as the sets

$$B_r(x) = \{y | d(x, y) < r\} \tag{1}$$

The strict inequality in equation (1) is the reason why these are considered “open" balls. Informally, when we refer to the closure of a ball $B_r(x)$, we are referring to the set of points such that $d(x, y) = r$.

**Def 2.5.** We define a simplex as the graph induced on a set by connecting every point in the set with a line. A simplicial complex is a set of simplices.

The concept of a simplex is best understood by considering simplices for sets of different dimensions. Let $S$ be an arbitrary set of points and $k = \dim(S)$ the dimension of $S$.

- If $k = 1$, the dimension of the structure created by connecting every element of $S$ is zero: a single point.
- If $k = 2$, the dimension of the structure created by connecting every element of $S$ is 1: a line connecting two points.
- If $k = 3$, the dimension of the structure created by connecting every element of $S$ is 2: two lines connecting three points and the face of the triangle created between them.

**Def 2.6.** Informally, homology groups are a means of encoding information about how many holes are present in a given space, where $H_n$ refers to the homology group of dimension $n$. $H_0$, the group of dimension $0$, encodes information about connected components; $H_1$ encodes information about loops; and $H_2$ encodes information about enclosed voids, or “three-dimensional loops." Simplicial homology refers to the homology groups of a simplicial complex.

**Def 2.7.** The Betti number of a given dimension for a simplicial complex represents the rank of the corresponding homology group. In particular, $b_0$ gives the number of connected components, $b_1$ the number of loops, and $b_2$ the number of voids.

Altogether, the simplicial complex of a set is a lower dimensional representation of the set that, remarkably, maintains the topological signatures of the original data set. This is summarized in the Nerve Theorem, which says that a simplicial complex constructed from a set $S$ is homotopically equivalent to $S$.

2.4. **The TDA pipeline.** We begin with the set $D$ of unordered points in three dimensions defined as

$$D = \{(x_0, x_1, x_2)_i\} = \{\mathbf{x_i}\} \in \mathbb{R}^3 \tag{2}$$

Note that if we attempt to create simplices on the set $S = D$, the entire subset of data, we are connecting every point in $D$ with an edge and creating a number of simplices on the same order of magnitude as the number of points in my data set. This doesn't serve us well if we are trying to deduce structure and see past noise in the data set.

To refine this approach and ensure that our simplicial complexes are actually of lower dimension than our initial data, we impose a filtration on the data. This filtration amounts to creating balls of radius $r$ centered at every point in the data set and adhering to the following rule: we can only connect two points if the closures of the open balls centered at those points intersect. The result is that we get very different simplicial complexes for different values of $r$. As we increase $r$, the resolution of our filtration changes and we gain and lose topological features. The persistence diagram gives us information about our data set that is significantly lower dimensional than the original set, but is still fairly high dimensional and therefore difficult to digest. One way to further distill the topological information we've uncovered about our data is to compute the Betti numbers for the corresponding simplicial complex. The Betti numbers give us a notion of the connectivity of our set in 1-,2- and 3-dimensions. Of course, these numbers will change as we change our radius $r$, and the way in which the Betti numbers evolve as we change our radius may give us an idea of the connectivity of our data set at different scales.

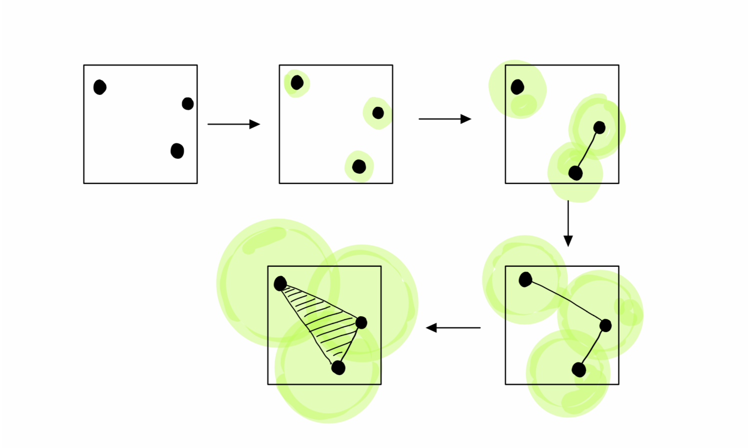


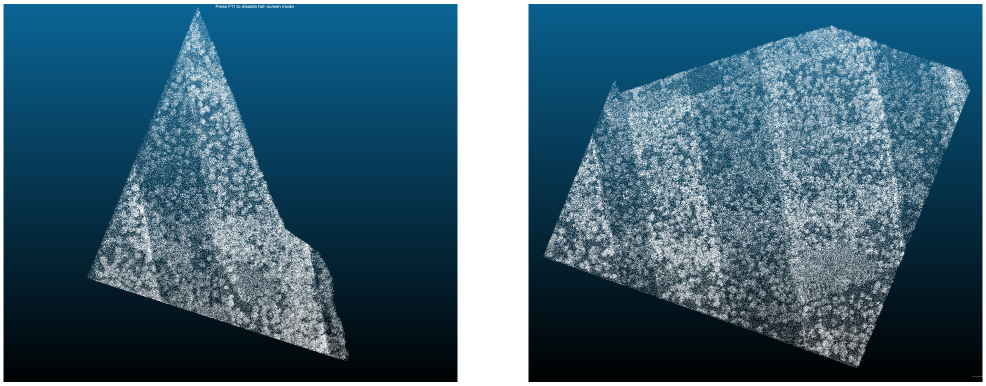
**FIGURE 2.** Visualizing filtration and the construction of simplices

2.5. **Choice of simplicial complex.** Simplicial complexes can be constructed from simplices in a number of ways resulting in an array of named complexes, each with its own distinct set of rules. The Vietoris-Ripps and Čech complexes are most commonly used. Other options include the witness complex, the cubical complex (used with voxelized data) and the alpha complex. Each has its own advantages and disadvantages, and the choice of complex is generally linked to the format and structure of one's data.

For this project, I began using the Vietoris-Ripps (VR) complex. I quickly ran in to processing and memory issues on my machine, despite thinning the original data to the point where it was no longer recognizable as fuel structure data. The problem lay in the fact that VR algorithms rely on combinatorial indexing to keep track of the growing number of simplicial complexes as the filtration radius changes. So, for $k$ data points and dimension $d$ features, we have

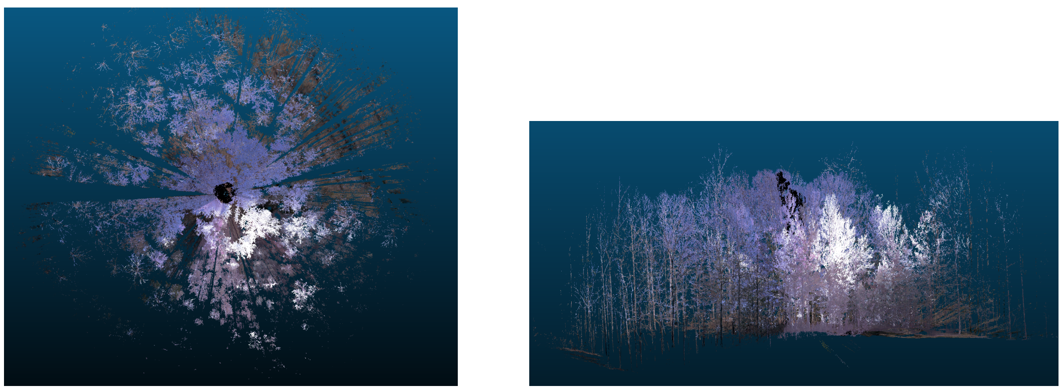
$$\binom{k}{d+1} = \frac{k!}{(d+1)!(k-d)!} \tag{3}$$

simplices. Clearly, the number of simplices grows rapidly with $k$, requiring a untenable amount of memory for larger point clouds. An alternative, the alpha complex, has dimension bounded by the dimension of the data set we're working with [6]. The alpha complex is a subset of the Delauney triangulation: see [7] for details. For computational reasons the alpha complex doesn't work well for higher-dimensional data. For our three-dimensional LiDAR data, however, the alpha complex works very well, allowing us to do all our processing on a personal laptop computer.



(A) Tile 1 (B) Tile 5

**FIGURE 3.** Example ALS scans



(A) Top view (B) Side view

**FIGURE 4.** Example BLK scan

2.6. **Data pre-processing.** Before processing with TDA algorithms, all data sets were normalized, thinned, elevation gradients were removed and, in some cases, the data were clipped to the area of interest. All pre-processing was done in R using the lidR package [8]. Visualization of the initial point cloud data was done using CloudCompare [9].

Our data fell in to three categories: full ALS tiles, BLK scans, and ALS tiles clipped to a 12 meter radius and centered at the center of the corresponding BLK scans. After thinning, the full ALS tiles were on the order of $500,000$ individual data points and the clipped ALS and BLK scans were both on the order of $50,000$ points. The BLK scans required the most severe thinning, with the original point clouds being on the order of $50$ million points. The un-thinned ALS tiles were on the order of $15$ million points, and the clipped ALS tiles were not thinned at all.

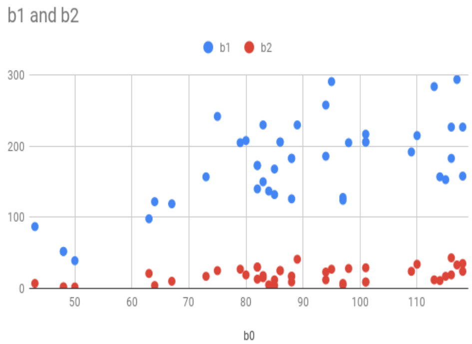
2.7. **Results.** Data was processed using the alpha complex class in the GUDHI library [10]. The homology of the resulting alpha complexes on the data was analyzed and the output was stored in a unique persistence diagram for each data set: see figure 6 for a visualization. These persistence diagrams encode information about the topological structure of our data, as previously discussed, but are themselves high dimensional objects that can be challenging to interpret. Nevertheless, the initial persistence diagram outputs point to clear topological differences between data sets. See, for example, the persistence diagrams in figure 7, which correspond to the ALS scans in figure 3.

Red dots indicate $0^{\text{th}}$ homology features, blue dots indicate $1^{\text{st}}$ homology features, and red dots indicate $2^{\text{nd}}$ homology features. Features clustered along the diagonal are born and die in a short period of time, whereas features far from the diagonal persist over a larger range of radii. Features present along the line $y = +\infty$ persist until the end of our filtration. Note that, for large $\alpha^2$ values, we have only a single connected component persisting to the end of our filtration.

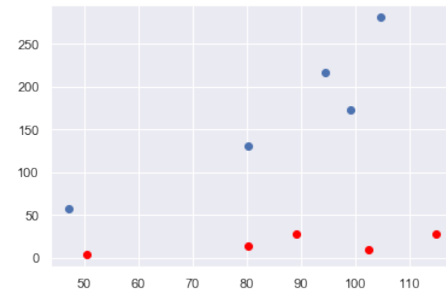
To further reduce the dimensionality of our data and understand our results, we calculated the Betti numbers over a range of $\alpha$ values. To hone in our analysis, we focused on the BLK scans. We observed that, for different plots, the Betti triple $[b_0, b_1, b_2]$ converged to $[1, 0, 0]$ at different rates, indicating varying levels of connectivity across scales in different plots. Moreover, when looking at $b_1$ and $b_2$ plotted as a function of $b_0$–a measure of the relative connectivity of the data set–we found distinct clustering emerging.

To make this observation about clustering concrete, we employed a k-means algorithm implemented in SciPy. The output of the k-means algorithm confirmed the clusters that we could intuitively "see" in the data. The observed clustering behavior hints at a means of classifying our BLK data according to its underlying topological structure, which is one of our three initially stated research objectives.

2.8. **Next Steps.** The immediate next step is to figure out what underlying characteristics of the data are driving this clustering behavior. This requires us to identify the appropriate number of clusters into which to group our relative Betti numbers. The choice of $k$ in $k$-means clustering is generally driven by knowledge of the underlying data, and experimenting with the clustering behavior of various data sets whose underlying characteristics are known should help us nail down the correct choice for $k$.

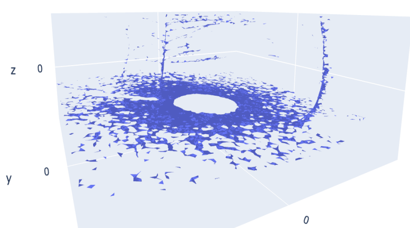



(A) all scans, $\alpha^2 = 5$

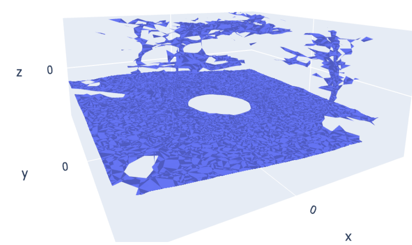



(B) centers of corresponding k-means clusters, $k = 5$

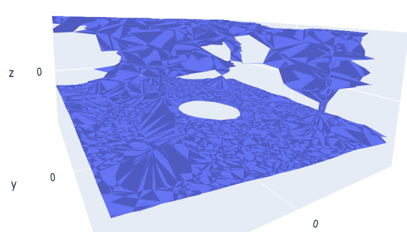
**FIGURE 5.** Clustering behavior in relative Betti numbers



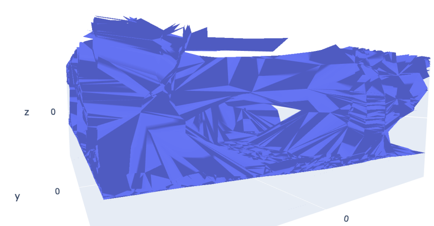
(A) $\alpha^2 = 0.005$



(B) $\alpha^2 = 0.05$



(C) $\alpha^2 = 0.5$

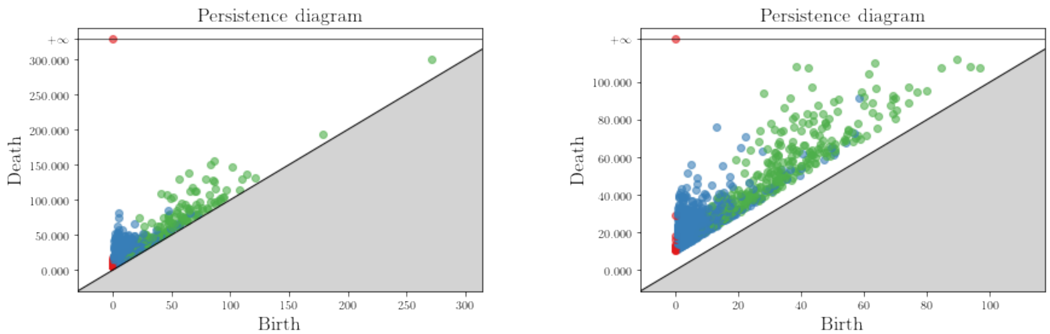


(D) $\alpha^2 = 2.5$

**FIGURE 6.** Vizualizing the alpha complex for various values of alpha squared

The question of identifying an appropriate choice for $k$ is indicative of a broader class of question: now that we've identified topological features in some of our data, how do we identify what features in our original data generate those topological features?

One approach is to identify the algebraic generators of the simplices that make up our simplicial complexes, in order to pinpoint the exact spatial points in our data set that correspond to the persistent topological features we observe. This simple proposition may prove computationally challenging.




(A) ALS Tile 1 (B) ALS Tile 5

**FIGURE 7.** Example Persistence Diagrams

Another idea is to analyze multiple scans of the same plot, particularly pre- and post-burn scans, in the hope that the difference in the resulting persistent homology may clarify its roots.

Further work needs to be done on processing the ALS tiles and examining their persistent homology and relative Betti characteristics. Clarification of the $\alpha^2$-value at which features are resolved in each of the three types of scans may clarify how topological features scale between terrestrial and aerial data, but this requires a deeper understanding of what it means to resolve features. For example, a filtration that captures the relative connectivity of the grasses and shrubs in a BLK scan will likely fail to capture the connectivity of the tree stems in the same scan, simply because these features are on different scales. An in-depth analysis of how choice of filtration radius relates to the scale of the structures of interest would be interesting and beneficial.

BLK scans are, by nature, of non-homogeneous density. In particular, the point density is much higher at the center of the plot close to the scanner. We're not yet sure if this artificially inconsistent density is affecting the results of our topological analysis, but one way to test this possibility would be to voxelize the data. Working with voxelized data instead of point cloud data requires a different TDA pipeline employing cubical complexes as the building blocks of our topological structure.

Lastly, relating our analysis of fuels structure back to fire behavior would be worthwhile. In particular, I am interested in studying the dynamic progression of a fire front across fuel domains exhibiting varying degrees of connectivity.

## 3. Contributions Made to the Research Project

The application of topological data analysis to LiDAR forestry data is a novel approach to understanding fuel structure. Over the course of this project, we framed the problem of understanding fuel structure across scale; identified how to adapt TDA methods to the available data, modify our methods to data at different scales, and pre-process data; worked at the intersection of ecology and mathematics to understand what topological features might signify in an ecological context; and applied machine learning and data visualization techniques to further probe our results. We made significant progress in addressing point (2) in our stated three-part research question and formulated strategies to address points (1) and (3). Lastly, through presentations to and discussions with various teams working at the Southern Research Station, I communicated the potential of topological data analysis to address a wide range of questions in fuels and fire science.

## 4. What new skills and knowledge did you gain?

In terms of hard skills, this internship significantly increased my proficiency at writing scripts in both Python and R, particularly in data mining and processing applications. I also gained valuable knowledge about the relationship between fire behavior and fuel structure, including ecology and fire physics concepts to which I had not previously been exposed.

More importantly, I learned how to better communicate with people outside of my field and frame their problems in terms of my area of expertise. The process of taking a question rooted in ecology, understanding it as a question about data, and then formulating a solution strategy in terms of mathematics was more difficult than I expected. As an applied mathematician, I anticipate I will be engaging in this process of scientific translation for the remainder of my career, and I am grateful for the opportunity to practice an invaluable skill.

## 5. Research Experience Impact on My Academic/Career Planning

Working with members of the broader fire science community gave me a much clearer idea of the kinds of questions being asked in fire science and the tools being used to answer those questions. I now understand the challenges the community is facing in the development of accurate modeling and simulation tools in much greater detail than would have been possible without the benefit of the many conversations with domain experts I had throughout the course of my internship. This detailed technical understanding helps me clarify how my own mathematical interests dovetail with current fire science research. I have three years remaining in my graduate program, and am at the point where I am ready to begin formulating my dissertation question. I look forward to doing so with the perspective I gained from this internship.

## 6. Relevance to the mission of NSF

The stated mission of the NSF includes: "To promote the progress of science" and "to advance the national health, prosperity and welfare." From a fundamental research perspective, this project helped to advance understanding of potential applications of topological data analysis, a new and growing branch of applied mathematics. Moreover, research in to fire behavior and prescribed fire best practices is imperative to address the detrimental effects of uncontrolled wildfires, which the Centers for Disease Control and Prevention identifies as a threat to public health [11].

## 7. Acknowledgements

My USFS mentor, Dr. Louise Loudermilk, provided background information, discussion, ideas, and connections with other researchers for the duration of my internship. Other members of the Center for Forest Disturbance Science who played an important role in this project include Dr. Wade Ross, who provided insight and help with data pre-processing; Derek Wallace, who provided much of the raw LiDAR data; Christie Hawley, who provided detailed field data; Dr. Joseph O'Brien, who leads the team at the Athens Prescribed Fire Laboratory; and Dr. Scott Goodrick, who shared many good ideas about relating this analysis to modeling needs. Collaborators from outside agencies include the QUIC-fire team at Tall Timbers Research Station and Dr. Rodman Linn of Los Alamos National Lab, who provided insight in to the workings of FIRETEC and QUIC-fire and the needs of the fire modeling community.

# REFERENCES


[1] E. Rowell, E. L. Loudermilk, C. Hawley, S. Pokswinski, C. Seielstad, L. Queen, J. J. O'Brien, A. T. Hudak, S. Goodrick, and J. K. Hiers, "Coupling terrestrial laser scanning with 3d fuel biomass sampling for advancing wildland fuels characterization." *Forest Ecology and Management*, vol. 462, pp. 117 945–, 2020.

[2] F. Pimont, R. Parsons, E. Rigolot, F. de Coligny, J.-L. Dupuy, P. Dreyfus, and R. R. Linn, "Modeling fuels and fire effects in 3d: Model description and applications," *Environmental Modelling I& Software*, vol. 80, pp. 225–244, 2016.

[3] R. Ottmar, "3d fuel characterization for evaluating physics-based fire behavior, fire effects, and smoke models on us department of defense military lands," funded proposal. [Online]. Available: https://www.serdp-estcp.org/Program-Areas/Resource-Conservation-and-Resiliency/Air-Quality/Fire-Emissions/RC19-1064/RC19-1064/(modified)/06May2019

[4] A. Hudak, "Object-based aggregation of fuel structures, physics-based fire behavior and self-organizing smoke plumes for improved fuel, fire, and smoke management on military lands," funded proposal. [Online]. Available: https://www.serdp-estcp.org/Program-Areas/Resource-Conservation-and-Resiliency/Air-Quality/Fire-Emissions/RC20-1346

[5] E. V. Brender, "A guide to the hitchiti forest research center," 1952.

[6] H. Edelsbrunner and E. P. Mucke, "Three-dimensional alpha shapes," *ACM Transactions on Graphics*, vol. 13, p. 43–72, 1994.

[7] K. Fischer, "Introduction to alpha shapes."

[8] Jean-Romain Roussel , "Lidr: Airborne lidar data manipulation and visualization for forestry applications." [Online]. Available: https://cran.r-project.org/web/packages/lidR/index.html

[9] Daniel Girardeau-Montaut, "Cloudcompare." [Online]. Available: http://www.cloudcompare.org/

[10] GUDHI, "Geometry understanding in higher dimensions." [Online]. Available: https://gudhi.inria.fr/

[11] C. for Disease Control and Prevention. (2020) Climate effects on health: Wildfires. [Online]. Available: https://cdc.gov/climateandhealth/effects/wildfires.html