

Using Contour Trees in the Analysis and Visualization of Radio Astronomy Data Cubes

Paul Rosen*

University of South Florida

Adam Ginsburg ¶

National Radio Astronomy Observatory

Bei Wang†

University of Utah

Anil Seth ‡

University of Utah

Betsy Mills §

San Jose State University

Jeff Kern **

National Radio Astronomy Observatory

Julia Kamenetzky ¶¶

Westminster College

Chris R. Johnson ††

University of Utah

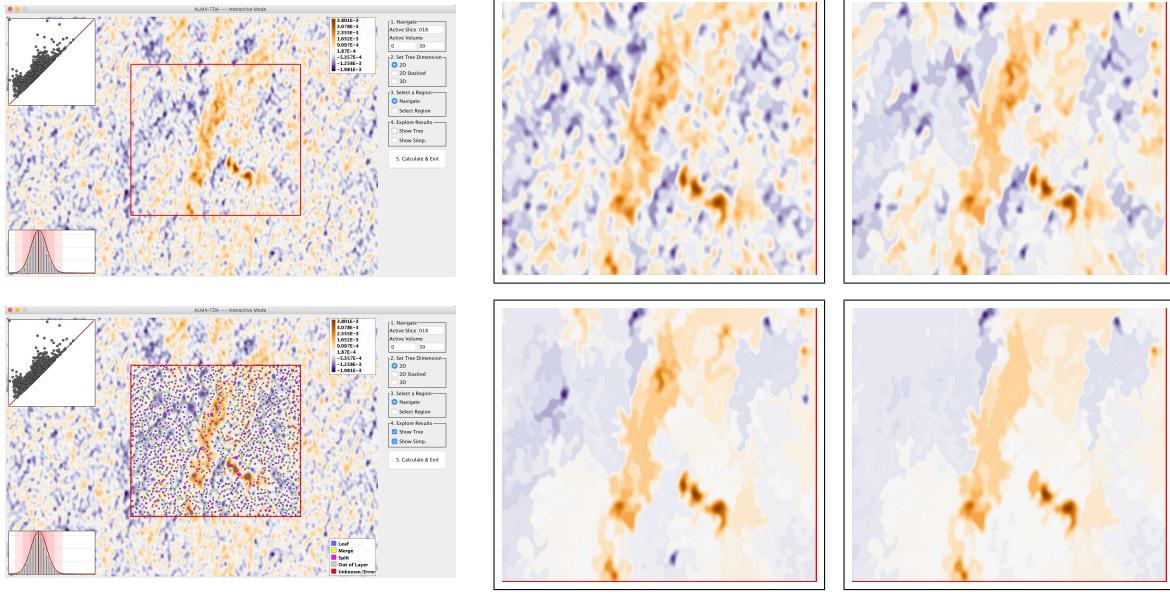


Figure 1: The results of varying the simplification level on slice #18 of the Ghost of Mirach dataset using a 2D contour tree. On the top left, a visualization of the original data is shown. On the bottom left, the contour tree computed on the region is shown with circles at critical point locations. Finally, on the right, the results of simplifying the data with simplification levels of 0.0005, 0.001, 0.0015, 0.002, from left to right, top to bottom, respectively.

ABSTRACT

The current generation of radio and millimeter telescopes, particularly the Atacama Large Millimeter Array (ALMA), offers enormous advances in observing capabilities. While these advances represent an unprecedented opportunity to advance scientific understanding, the increased complexity in the spatial and spectral structure of even a single spectral line is hard to interpret. The complexity present in current ALMA data cubes therefore challenges not only the existing tools for fundamental analysis of these datasets, but also users' ability to explore and visualize their data. We have performed a feasibility study for applying forms of topological data analysis and

visualization never before tested by the ALMA community. Through contour tree-based data analysis, we seek to improve upon existing data cube analysis and visualization workflows, in the forms of improved accuracy and speed in extracting features. In this paper, we review our design process in building effective analysis and visualization capabilities for the astrophysicist users. We summarize effective design practices, in particular, we identify domain-specific needs of simplicity, integrability and reproducibility, in order to best target and service the large astrophysics community.

1 INTRODUCTION

Radio astronomy is currently undergoing a revolution driven by new high spatial and spectral resolution observing capabilities. The current generation of radio and millimeter telescopes, particularly the Atacama Large Millimeter Array (ALMA), offers enormous advances in capabilities, including significantly increased sensitivity, resolution, and spectral bandwidth. While these advances represent an unprecedented opportunity to advance scientific understanding, they also pose a significant challenge. In some cases, the higher sensitivity and resolution they provide yield new detections of sources with well-ordered structure that is easy to interpret using current tools (e.g., HL Tau [41]). However, these advances equally often lead to the detection of structure with increased spatial and spectral complexity (e.g., new molecules in the chemically-rich massive star

*e-mail: prosen@usf.edu

†e-mail: beiwang@sci.utah.edu

‡e-mail: aseth@astro.utah.edu

§e-mail: elisabeth.mills@sjsu.edu

¶e-mail: aginsbur@aoc.nrao.edu

¶¶e-mail: jkamenetzky@westminstercollege.edu

**e-mail: jkern@nrao.edu

††e-mail: crj@sci.utah.edu

forming region Sgr B2, outflows in the nuclear region of the nearby galaxy NGC 253, and rich kinematic structure in the giant molecular cloud “The Brick” [8, 10, 47]). The new complexity present in current spectral line datasets challenge not only the existing tools for fundamental analysis of these datasets, but also users’ ability to explore and visualize their data.

1.1 Contribution

Inspired by Topological Data Analysis (TDA), we perform a *feasibility study* for applying forms of data analysis and visualization never before tested by the ALMA community. Through contour tree-based techniques, we seek to improve upon existing data cube analysis and visualization, in the forms of improved accuracy and speed in finding and extracting features.

We present the resulting software, called *ALMA-TDA* (<https://github.com/SCIIInstitute/ALMA-TDA>), which uses contour trees to extract and simplify the complex signals from noisy radio astronomy data, commonly referred to as *ALMA data cubes*. ALMA is an astronomical interferometer of radio telescopes located in northern Chile; it delivers data cubes (sometimes referred to as position-position-frequency cubes), of which the first two axes are positions and the third axis is the frequency or spectral domain.

An example of our tool is seen in Figure 1. In this figure, the original noisy data is shown on the upper left. The lower left shows the critical points extracted by the contour tree. Finally, the four images to the right show differing levels of simplification of the data.

In this paper, we review our design process in building effective analysis and visualization capabilities for ALMA data cubes. We summarize effective design practices targeting and servicing the large astrophysics community, in particular, we need to design tools with *simplicity* (i.e., light-weight), *integrability* (i.e., integrable within existing tool chains) and *reproducibility* (i.e., fully recorded analysis history via command-lines). We hope such learned design practices will provide guidelines toward future development of tools and techniques that would benefit astrophysicists’ scientific goals.

2 SCIENCE CASE

The new complexity brought about by increased sensitivity, spatial and spectral resolution, and spectral bandwidth has become a significant bottleneck in science, as it not only challenges astrophysicists’ analysis tools but also the users’ ability to understand their data. For example, increased complexity in the spatial and velocity structure of spectral line emission makes even a single spectral line hard to interpret. When a cube contains the superposition of multiple spatial and kinematic structures, such as outflows and rotation and infall, each with their own relationship between the observed velocity and their actual position along the line of sight, traditional analysis and exploration tools fail. Users struggle to follow kinematic trends across multiple structures by examining movies or channel maps of the data. However, moment map analysis (e.g., integrated fluxes, mean velocities and mean line-widths), the most commonly used analysis tool for compressing this 3D information into a more easily parsed 2D form, no longer has a straightforward interpretation in the presence of such complex structure, in which mean velocities may be velocities at which no emission is actually present, and mean line widths may represent the distance between two velocity components, rather than the width of a single component.

Whether scientists can navigate and correctly interpret this new complexity will determine their success in addressing a number of important scientific questions. Among the topics driven by the detection of more complex structures are ISM turbulence [21, 48], the star formation process [31], filaments [45], molecular cloud structure and kinematics [47], and the kinematics of nearby galaxies [29, 30, 34] and high redshift galaxies [42, 55].

An even greater challenge arises from our ability to detect an increased number of spectral lines in more and more sources. There

simply are no tools capable of simultaneously visualizing, comparing, and analyzing the dozens to hundreds of data cubes for all of the detected spectral lines in a given source. Such standard methods of both visualizing and exploring data as moment maps, channel maps, playing cube like a video, or 3D models, cannot scale up to the case of large numbers of lines, even in non-complex, well-ordered cases, such as rotating disks, or expanding stellar shells. Users become overwhelmed by, for example, comparing these typical diagnostics for two lines, side by side or one at a time. In the richest sources with thousands of lines, such comparisons will simply be impossible—it becomes necessary to resort to methods that entirely throw away either the spectral information such as Principle Component Analysis (PCA) of moment maps or the spatial information that requires model fitting of complex spectra. As a result, both exploration and analysis of the data becomes not only time consuming, but potentially incomplete.

As we move into the future and these telescopes reach their full potential, complex spatial and velocity structure will no longer be a problem that typically occurs in a separate subset of sources than those exhibiting rich spectra—the two problems will coexist, compounding the highlighted issues. The visualization and analysis challenges currently facing radio astronomy will then only grow more pressing as instruments grow more sensitive and the data volumes become larger.

2.1 Existing Tools

A critical aspect to the study of ALMA data cubes is the detection, extraction and characterization of objects such as stars, galaxies, and blackholes. *Source finding* in radio astronomy is the process of detecting and characterizing objects in radio images (in the forms of data cubes), and returning a survey catalogue of the extracted objects and their properties [27, 56]. A common practice is to use a computer program (i.e., a source finder) to search the data cubes, followed by manual inspection to confirm the sources of electromagnetic radiation [56]. An ideal source finder aims to determine the location and properties of these astronomical objects in a complete and reliable fashion [27]; while manual inspection is often time-consuming and expensive.

Several existing tools have been used in the ALMA community in terms of source finding [28], including the popular ones such as clumpfind [57], dendograms [51], crops [50], and more recent ones such as FellWalker [9], SCIMES [20] and NeuroScope [35]. Clumpfind is designed for analyzing radio observations of molecular clouds obtained as 3D data cubes, it works by contouring the data, searching for local peaks of emission and following them down to lower intensity levels [57]. The dendograms of a data cube is an abstraction of the changing topology of the isosurfaces as a function of contour level, which captures the essential features of the hierarchical structure of the isosurfaces [51]. The FellWalker algorithm is a gradient-tracing watershed algorithm that segment images into regions surrounding local maxima [9]. FellWalker provides some ability to merging clumps, therefore simplify the underlying structures, and the merging criteria shares some similarities with persistence-based simplification. However these criteria are less mathematical rigorous compared to our approach. SCIMES (Spectral Clustering for Interstellar Molecular Emission Segmentation) considers the dendrogram of emission under graph theory and utilizes spectral clustering to find discrete regions with similar emission properties [20]. Finally, the most recent NeuroScope [35] (specifically targeted for ALMA data cubes) employs a set of neural machine learning tools for the identification and visualization of spatial regions with distinct patterns of motion.

However, the study of source finding for ALMA data cubes raises the following question: How can we help the astrophysicists to understand the *de-noising* process? That is, how to best separate signals from noise, and the effects of de-noising on the original data?

In other words, it is important for us to *quantify both signals and noise* as well as to *perform simplifications* of the underline data. This kind of study is underdeveloped with current approaches in the ALMA community .

3 TECHNICAL BACKGROUND: CONTOUR TREES

From a technical perspective, we focus on performing data analysis and designing effective visualization of ALMA data cubes by employing the contour tree [13]. The contour tree is a mathematical object describing the evolution of the level sets of a scalar function defined on a simple, connected domain, such as the grayscale intensity defined on the 2D domain associated with a slice of a data cube (at a fixed frequency). There are two key properties associated with a contour tree, making it a feasible tool in the study of ALMA data cubes. First, a contour tree has a graph-based representation that captures the changes within the topology of a scalar function and provides a meaningful summarization of the associated data. Second, a contour tree can be easily simplified, in a quantifiable way, to remove noise while retaining important features in data.

3.1 Contour Tree

Scalar functions are ubiquitous in modeling scientific information. Topological structures, such as contour trees, are commonly utilized to provide compact and abstract representations for these functions. The contour tree of a scalar function $f : \mathbb{X} \rightarrow \mathbb{R}$ describes the connectivity of its *level sets* (isosurfaces) $f^{-1}(a)$ (for some $a \in \mathbb{R}$), whose connected components are referred to as *contours*. Given a scalar function defined within some Euclidean domain \mathbb{X} , the contour tree is constructed by collapsing the connected components of each level set to a point. The contour tree stores information regarding the number of components at any function value (isovalue) as well as how these components split and merge as the function value changes. Such an abstraction offers a global summary of the topology of the level sets and enables the development of compact and effective methods for modeling and visualizing scientific data. See Figure 2(a)-(c) for an illustrative example. Vertices in the contour tree correspond to *critical points* of the 2D scalar function, namely, local minima, saddle points, and local maxima, whose local structures are illustrated in Figure 3.

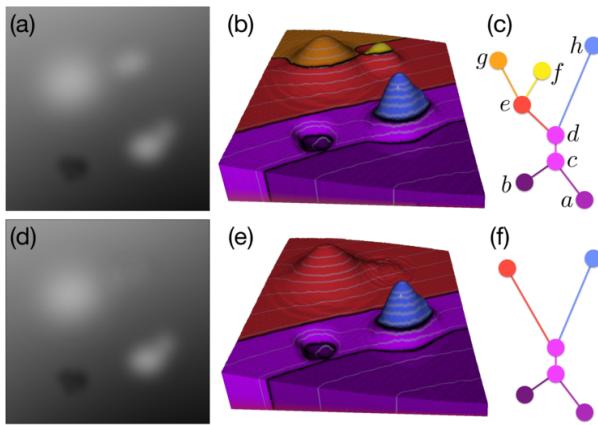


Figure 2: (a) A grayscale image of a 2D scalar function before simplification. (b) 3D height map of the contours corresponding to the scalar function shown in (a). (c) The contour tree structures that capture the evolution of terrain features (i.e., relations among local minima, local maxima, and saddle points). (d)-(f): The grayscale image, 3D height map, and the contour tree after simplifying the terrain features.



Figure 3: Local structures of critical points. From left to right: a local minimum, a saddle point, and a local maximum.

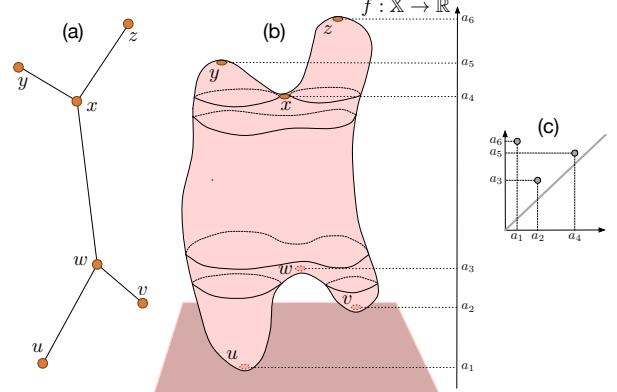


Figure 4: Example of persistence pairing of critical points for (b) a 2D height function. The persistence pairing of (a) critical points from the contour tree gives rise to (c) a (scaled) persistence diagram.

3.2 Contour Tree and Scalar Field Simplification

To simplify a contour tree, we assign an importance measure to each edge of the tree and collapse (eliminate) edges of lower importance measures [12, 14]. Various geometric properties, such as persistence, volume, and surface area, can be used to compute the importance measure. We apply ideas from topological persistence [22] in our feasibility study.

Persistence. We now describe the idea of persistence in a nutshell [4], using Figure 4 as an illustrative example. Given a height function $f : \mathbb{X} \rightarrow \mathbb{R}$ defined on a 2D domain, let \mathbb{X}_a denote the sublevel set of f , that is, $\mathbb{X}_a = f^{-1}(-\infty, a]$. Suppose we sweep a horizontal plane in the direction of increasing height values, and keep track of the (connected) components of \mathbb{X}_a while increasing a . A component of \mathbb{X}_a starts at a local minimum, and ends at a (negative) saddle point when it merges with an older component (i.e., a component that starts earlier). This defines a minimum-saddle persistence pair between critical vertices, and the *persistence* of such a pair is the height difference between them. Similarly, a hole of \mathbb{X}_a starts at a (positive) saddle point and ends at a local maxima (where it is capped off). This defines a saddle-maximum pair with its persistence being the height difference between its vertices.

Referring to Figure 4: points u and v are local minima; y and z are local maxima; w is a negative saddle point; and x is a positive saddle point. Their corresponding height values are sorted as $a_1 < a_2 < \dots < a_6$. We sweep a horizontal plane in the direction of increasing height value and keep track of the components in the sublevel set. The pair (v, w) forms a minimum-saddle persistence pair, as a component in the sublevel set starts at point v and it merges with an older component that starts at point u . The pair (x, y) form a saddle-maximum pair, as a hole starts at x and ends at y . The pair (v, w) has a persistence of $|a_2 - a_3|$; while the pair (x, y) has a persistence $|a_5 - a_4|$. The contour tree is shown in Figure 4 (a).

Persistence Diagram. The pairing of critical points also give rise to a *persistence diagram* [18] that summarizes and visualizes topological features of a given function. A persistence diagram contains a multiset of points in the plane; its x - and y -coordinates captures the start (*birth*) time and the end (*death*) time of a particular topological feature. The distance of the point to the diagonal captures the persistence of that feature. Points away from the diagonal have high persistence, and correspond to signals of the data; while points that are close to the diagonal have low persistence, which are typically treated as noise.

In the example of Figure 4, the critical point pairs (x, y) and (v, w) give rise to points (a_4, a_5) and (a_2, a_3) in the persistence diagram, respectively. This persistence diagram also contains an additional off-diagonal point (a_1, a_6) , which corresponds to the pairing of global minimum u with the global maximum z that captures the entire shape of data. This is a global feature that can not be simplified (see [19] for technical details). In our context, we only care about minimum-saddle and saddle-maximum pairs.

Contour Tree Simplification. In the contour tree example of Figure 2(c), the pair (b, c) is a minimum-saddle pair while the pairs (e, f) and (d, g) are saddle-maximum pairs. During a hierarchical simplification, the pair (e, f) has the smallest persistence, therefore the edge connecting them is collapsed (simplified), as shown in Figure 2(f); this can be achieved by a smooth deformation of the surface in Figure 2(d). This deformation is detailed below.

Scalar Field Simplification. Given a contour tree simplification, we would like to compute its corresponding scalar field simplification. Simplifying a scalar function directly in a way that removes topological noise as determined by its persistence diagram has been investigated extensively (e.g. [23]). As pointed by Carr et al. [15], contour tree simplification have well-defined effects on the underlying scalar field: collapsing a leaf corresponds to leveling off (or flattening) regions surrounding a maximum or a minimum. This is a desirable simplification for the domain scientists, as they are interested in reducing noise to zero flux during the de-noising process. Figure 2(b) and (c) demonstrate the result of edge collapsing: collapsing the edge (e, f) from the tree results in flattening the yellow region surrounding the local maximum f ; this is equivalent to introducing a small perturbation to the neighborhoods of saddle-maximum pair (e, f) so that both critical points e and f are removed. Such a flattening process is further highlighted in Figure 5, and in practice, it is implemented under the piecewise linear setting.

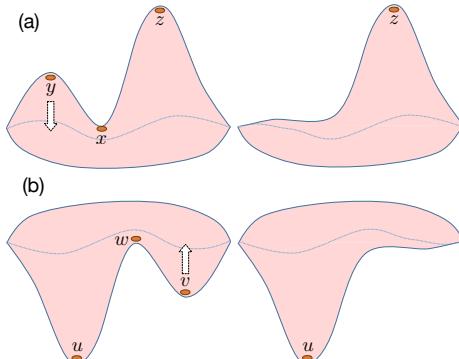


Figure 5: Simplifying a saddle-maximum pair (x, y) in (a) and minimum-saddle pair (v, w) in (b) for a 2D scalar field. (a): We reduce the height of the local maximum y to the level of saddle x , effectively flattening the region surrounding y . (b): We increase the height of local minimum v to the level of the saddle w , again flattening the region surrounding v .

3.3 Related Work

The contour tree was first introduced by van Kreveld et al. [54] to study contours on topographic terrain maps (i.e., curves containing sampled points with the same elevation values). It has then been widely used for both scientific and medical visualizations [5, 43, 52, 53]. Efficient algorithms for computing the contour tree [13, 17, 46] (and its variants, merge tree [39], and Reeb graph [44]) in arbitrary dimensions have been developed. The latest state-of-the-art regarding contour trees have been parallel or distributed implementations [11, 16, 26, 32, 36, 37]. We use an approach described in [49].

A related concept called dendograms has been used in astronomical applications to segment data and to quantify hierarchical structure such as the amount of emission or turbulence at a given size scale, for example, to study the role of self-gravity in star formation [25]. A dendrogram is a tree-diagram typically used to characterize the arrangement of clusters produced by hierarchical clustering. It tracks how components (clusters) of the level sets merge as the function value changes, while a contour tree captures more complete topological changes (i.e., merge and split) of the level sets. The state-of-the-art Astronomical Dendrogram method [1] has limited capabilities in automatic data denoising, feature extraction and interactive visualization.

4 DESIGN PROCESS

In this section, we revisit our design process in building effective analysis and visualization capabilities of ALMA data cubes. By reviewing the timeline of our project, we reflect on key design activities with the goal of learning from experience and summarizing effective design practices targeting and serving the astrophysics community.

Critical Activities. To give an overview of our design process, we describe the critical activities as identified in [33]: understand, ideate, make, and deploy.

The discussion of our project started at around November 2014, when National Radio Astronomy Observatory (NRAO) scientist Dr. Jeff Kern visited the Scientific Computing and Imaging (SCI) Institute and saw a talk on the topic of topological data analysis. Over the following months, follow-up conversations with the computer scientists on our team generated some initial excitement regarding the potential of applying topological techniques in understanding ALMA data cubes, which has never been done before.

To *understand* the problem domain and target users, we identified key opportunities, that is, applying emerging techniques from topological data analysis to the study of ALMA data cubes. Our main motivation stemmed from Jeff's comments that "there simply are no tools capable of simultaneously visualizing, comparing, and analyzing the dozens to hundreds of data cubes for all of the detected spectral lines in a given source." We believed that introducing topological data analysis techniques to the ALMA community would potentially offer new insights regarding feature detection, as well as improve their workflow efficiency.

The *ideate* activity of our project started in May 2015, as the domain problems became better characterized and possible solutions via contour tree-based approaches appeared to have the greatest potential among the solution space. Members in our team started working on a 1-year proposal under ALMA Development Studies regarding feature extraction and visualization of ALMA data cubes through contour tree centered approaches. The proposed work was subsequently funded. Writing a proposal targeting ALMA community helped us greatly in terms of externalizing our ideas and expected technical challenges, while at the same time, formulating a potential analysis pipeline, visual encodings, and selecting interactive capabilities within the proposed system.

By the time our project officially started in January 2016, computer scientists on our team have already met with astrophysicists at NRAO facility to learn their needs and conducted an on-campus

interview with astrophysicist, Dr. Anil Seth, who works with ALMA data cubes. We learned the typical pipeline in the analysis and visualization of ALMA data cubes, specifically, in Anil’s case, via image editing tools or file viewers such as QFitsView [2] and SAOImage DS9 [3]. We also gave short tutorials regarding our proposed techniques to obtain comments and feedback from all our interactions.

We started our *make* activity by constructing a tangible prototype, specifically encompassing visualization decisions and interaction techniques. Our process coupled the ideate and make activities in the design and refinement of our system. We identified that *quantification* (of signals and noise) and *simplification* are two of the most important aspects for our proposed framework. We went through multiple rounds of interface mock-ups and functionality discussions. We showcased our first prototype between June and August 2016 to the domain experts, including one-on-one discussions with Anil and Dr. Julia Kamenetzky, and through a number of talks given to the astrophysics community, with general positive feedback.

Over the course of the next half an year, we rolled out multiple phases of *deploy* activities, in order to put the prototype in real-world setting to understand how to improve its effectiveness and performance. Our goal was to have a usable system that helps with the users’ data-specific tasks.

In January 2017, we organized a one-day workshop with domain scientists, Julia, Dr. Betsy Mills, and Dr. Adam Ginsberg. We engaged in panel discussions on the current version of the prototype, gathered comments and suggestions, and discussed potential research and developmental directions moving forward. This workshop in particular helped to cement the lessons learned.

Effective Design Practices. Throughout our design process, we have learned a few effective design practices that best serve the ALMA community: simplicity, integrability, and reproducibility.

In terms of *simplicity*, the tool should contain sufficient but not overwhelming amount of visualization; and minimize GUI interactions. This philosophy is in sharp contract with some of our common practices as visualizers, where we aim to create novel, exciting and sometimes flashy visualizations. Our initial prototypes were full of many unnecessary tools and functionalities with complex GUIs. We learned via feedbacks and user practices that a complex interface will distract or confuse the users to the point that they won’t even try using the software. The tool should also be light-weight. That is, it should be easily installed on a desktop computer and not require external dependencies or packages be installed. For this reason, we chose Java as a platform from the beginning. Though not the most efficient, Java software is highly portable. This is well-aligned with properties of commonly used processing tools in the ALMA community.

In terms of *integrability*, the tool should be integrable with existing workflows and toolchains. This means that the core functionality of the software need to be automatable. In addition to providing a GUI, we also provide a command-line interface for generating results. One of our future directions is to link our tool directly with CASA, the Common Astronomy Software Applications package. CASA is under active development by NRAO and primarily supports the data post-processing needs of the next generation of radio astronomical telescopes such as ALMA and the Very Large Array (VLA).

In terms of *reproducibility*, the analysis history using our tool should be recorded so that the results can be reproduced. This is supported in two ways. First, by enabling processing via the command-line, we can save parameters and automatically rerun the results later. Second, we minimize the amount of GUI interactions, as most of such interactions are exploratory and do not necessarily lead to insights. When the user is satisfied with their results using our visualization, the exact command required to reproduce the results is output to the command-line for reference.

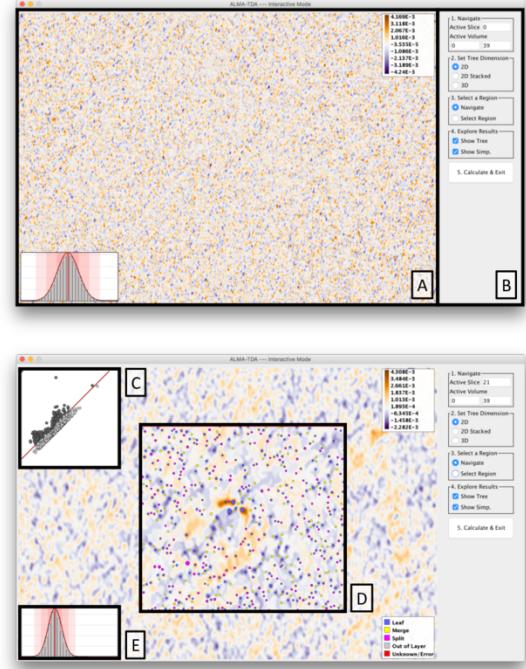


Figure 6: Upon loading the software in the interactive mode, the user is presented with a view of the data. Top: Initial view of the software. Bottom: Visualizations shown as the user selects a region of interest and the contour tree is calculated (on the back end).

5 SOFTWARE DESIGN

By default our software is a command-line tool. The command-line interface provides a small set of options for reproducibility of any computation. No other options are needed to reproduce results. Those options are:

- **Input file** – Path to the file for processing.
- **X, Y, & Z range** – The dimensions of the region to process.
- **Simplification type** – Either 2D, for a single slice; 2D Stack, for a series of 2D slices; or 3D, for volume processing.
- **Simplification level** – Persistence level for feature simplification.
- **Output file** – Path to save results.

We also provide an interactive mode to explore the capabilities of our approach and select these parameters. When starting the software, users need only add the “interactive” tag to the command-line, and the visualization launches.

The visualization initially opens to the interface seen in Figure 6 (top). The interface is designed to include only minimal required capabilities. The main window, A, shows visualizations related to the loaded data cube. The GUI component, B, provides controls to set options for processing the data. The controls are placed in groups, numbered for steps 1-5. The GUI component is designed with both simplicity and functionality in mind, to offer the users most intuitive, and yet fully-functional analysis capabilities.

5.1 Visual Elements

The visualization is composed of five main visual elements.

Scalar Field View (Figure 6 A). Being a sampling of radio waves, the 2D scalar field (a slice of ALMA data cube along the frequency axis) has both positive and negative amplitudes. It is

therefore displayed using a divergent orange/purple colormap. By default, the first slice is selected and viewed by centering on the middle of the domain. The user can translate and zoom with the mouse movement. Different slices can be selected by changing the values in the controls to the right.

Persistence Diagram (Figure 6 C). Once the contour tree is calculated, the data are displayed using a persistence diagram. A persistence diagram is a scatterplot display that places points representing features by their birth time horizontally and their death time vertically. A convenient feature of this chart is that distance from the diagonal is also the persistence of that feature. Therefore, we use this visualization for interactively selecting the level of simplification. As the red simplification bar is dragged, features below the bar are grayed out indicating that they will be canceled (simplified). Once released, a simplified contour tree (on the back-end) and a simplified scalar field (for the front-end) are calculated.

Contour Tree (Figure 6 D). Displaying the tree structure of the contour tree is not particularly meaningful, as it is both large and an abstract view of the data. However, seeing critical points and their persistence in the context of the data is valuable. The critical points are placed over the scalar field view at their respective spatial location. Their size is set based upon their persistence (higher persistence, larger point). Finally, their color is set by their type: local extrema (leaf of the tree) – blue, negative saddle points (merge) – yellow, and positive saddle points (split) – magenta. For 3D analysis, contour tree nodes off layer are colored gray. An error state, red, is provided for nodes that don't cancel. This would be caused by some kind of boundary effect that we have yet to investigate. This view of the contour tree can be enabled or disabled on demand using the controls on the right.

Simplified Scalar Field (Figure 6 D). Since users are in large part interested in the feature extraction power of this approach, we show the result of scalar field simplification in context. As the user adjusts the level of persistent simplification, the scalar field is simplified and overlaid with the original visualization. This view can be enabled or disabled on demand using the controls on the right.

Histogram (Figure 6 E). A histogram is produced, indicating the distribution of (intensity) values of data cubes within the current view. In addition to showing histogram bins, this view shows the mean as a solid red line and ± 3 standard deviations as consecutively lighter red bars. This histogram is adapted as the user navigates their view or when the simplification level of the scalar field is adjusted. This view is important, as domain experts are interested in quantifying the total flux gained or lost during simplification. This is most observable by shifts in the mean.

5.2 Interaction Process

Though the use of our software requires some explanation, we strive to make it as simple to use as possible. Part of this effort is providing a simple and intuitive five step approach to the users.

Step 1: Navigation. The users are first asked to navigate the view to the general region of interest. This includes translation and zooming, but it also includes selecting the slice or volume of interest.

Step 2: Tree Dimension. The dimension of the contour tree calculation must be selected next. The options include 2D, for a single slice; 2D stack, for 2D computation on a series of slices; and 3D for computation on a volume. These options will be discussed further in the case study.

Step 3: Region Selection. Next the specific region of interest is selected with the mouse. As soon as the mouse is released, computation begins. If the region is large, the user is prompted with the option to cancel, due to computation time. We are actively investigating scalable contour tree computations to support larger data cubes with on-the-fly visualization.

Step 4: Exploration. Once the computation is completed, the user is invited to explore the domain. This includes navigation (translation, zooming, and changing slices) and adjustment of the simplification level. As simplification levels are adjusted, the user can observe changes in the scalar field, compare those changes to the original field, and look for changes in flux in the histogram.

Step 5: Compute and Exit. Steps 1-4 may be repeated as many times as necessary, until the user is satisfied. Once done, the user clicks “Compute and Exit”. This will trigger a processing of the data cube and saving of output. Finally, the precise command required to reproduce the results will be printed on the command-line.

6 CASE STUDIES

We show the capabilities of our prototype with two case studies involving specific ALMA data cubes used by our collaborators.

6.1 Ghost of Mirach Galaxy Data Set

NGC 404 (also known as Mirach's Ghost) is data of a molecular gas emission at the center of the nearby, low mass galaxy. The data was taken using ALMA on Oct. 31st, 2015. A data cube was created using the default ALMA pipeline and involves Fourier transformation of the interferometric data at each frequency. The data cube is approximate 4.5GB with resolution of 5400x5400 in the spatial domain and 30 in the spectral domain (i.e., 30 slices). However, the feature of interest is around 200x200 in size and covers around 10 slices. Scientists often sample cubes much larger than their feature of interest to reduce some structured errors, vignetting for example.

Science Description. Excited molecular carbon monoxide gas emits light at 230 GHz. The doppler shifts of this line emission can provide information on the motion of molecular gas in the galaxy. Visualization of the data of NGC404 shows a clear rotating disk located within the central 20 light years of the galaxy. Similar rotating molecular gas disks have been used to measure the masses of supermassive black holes at the centers of galaxies (e.g. [6, 40]). However, the data is noisy, so coherent gas structures are hard to pick out. NGC 404 presents a special challenge due to the low mass of its black hole [38]. Fortunately, the high angular resolution of

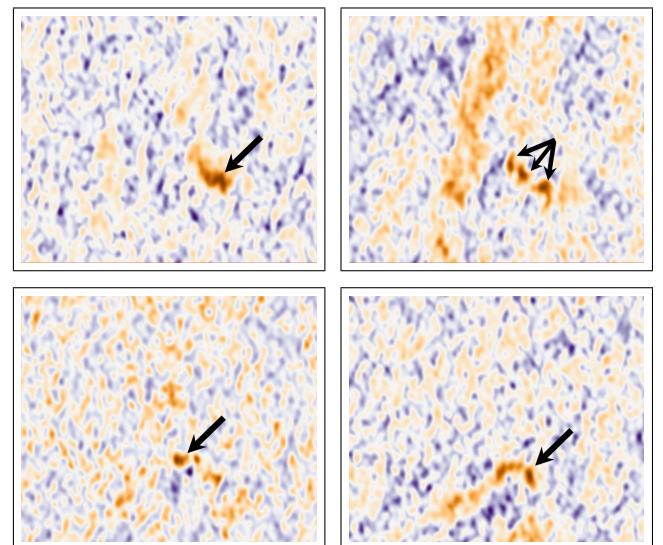


Figure 7: Four slices, #16, 18, 20, & 22 from the Ghost of Mirach dataset. The bright red spots (indicated by the arrows) in these images are the signal of interest.

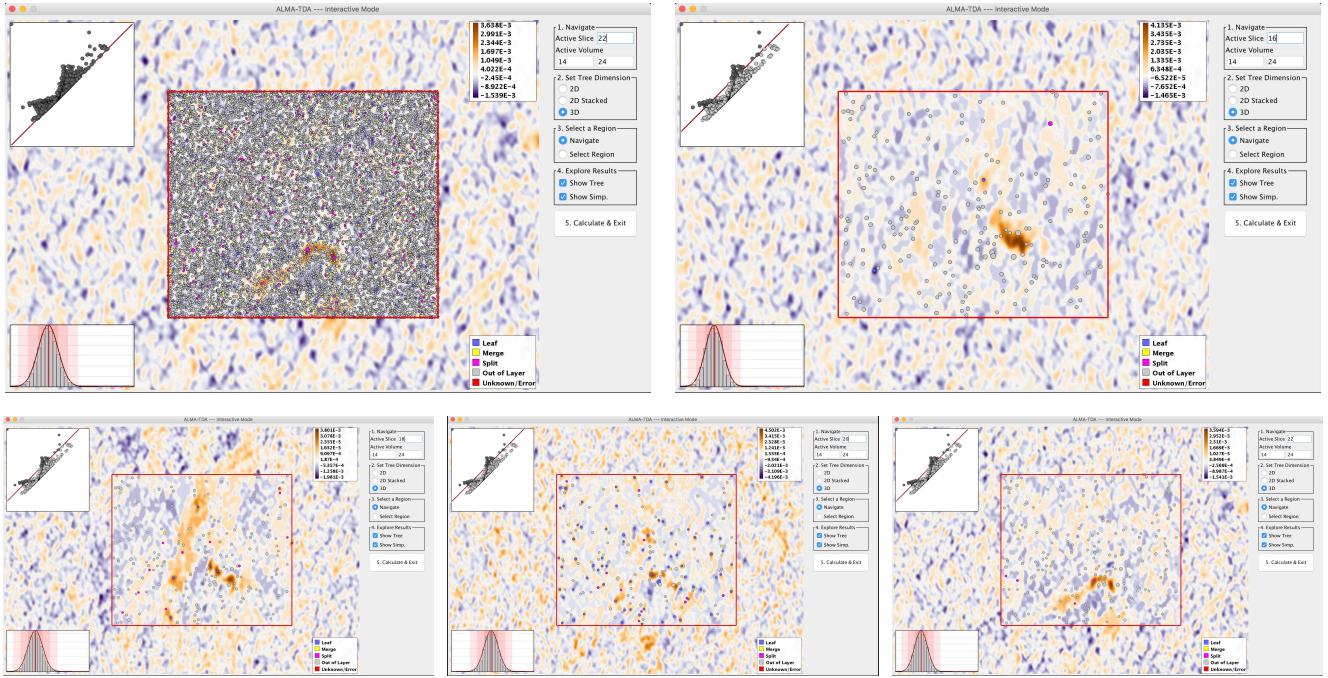


Figure 8: Result of simplifying using the 3D contour tree on the Ghost of Mirach dataset was worse than expect due to topological pants (tubes connecting through slices). Top left: Visualization of the 3D contour tree on slice 22. Top right: Simplification of slice 16. Bottom: Simplification of slices 18, 20, & 22, respectively. The persistent simplification level was 0.00128.

ALMA provides the highest sensitivity to measuring the black hole mass.

We can see an example of 4 spectral slices of the dataset in Figure 7. In these 4 slices, the bright red spots represents the signal, while most of the remaining patterns represent noise.

Varying Simplification Levels. Figure 1 shows an example of performing simplification on a single 2D spectra (i.e., a single slice along the frequency axis). The noisy structure is captured by the 2D contour tree as many low persistence features (lower left). Increasing the level of simplification removes much of this noise (lower middle). However, selecting a simplification level that is too aggressive may result in loss of signal (lower right).

3D Contour Trees. Since the spectral data are treated as cubes, our collaborators were interested in the structures that would be found using 3D contour trees. The result of capturing the 3D contour tree, shown in Figure 8, was both a surprise and a disappointment. Although many critical points were found, the data suffered from what we describe as *topological pants*. A *topological pair of pants* is a sphere with three disjoint closed discs removed [7]. Essentially, the 3D contours of noisy features form a complex interconnect tubes through the volume that are not physically meaningful. This interfered with the kind of features that a contour tree can identify. The root cause of this is that each of the spectra are processed independently, and thus, there is no correlation between noise patterns across consecutive slices. Simplifying these temporal noise patterns as a whole is not physically meaningful, and they interfere with true features in the data.

2D Contour Tree Stacks. On the other hand, the processing of 2D contour trees was highly successful. However, the domain scientists still needed the ability to process 3D cubes. The obvious solution was to use a series of 2D contour trees to control the simplification. Figure 10 shows the result of simplifying a stack of spectra. This example uses a similar level of simplification to the 3D contour tree example in Figure 8. In our implementation, level of

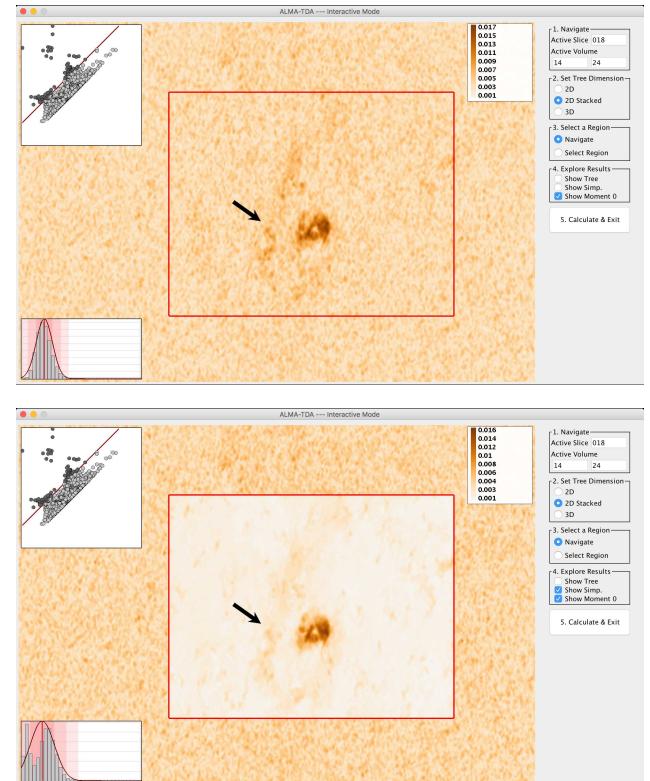


Figure 9: Moment 0 analysis of Ghost of Mirach dataset between slices 14 and 24 (the range of the signal) using a stack of 2D contour trees. Top: Visualization of moment 0 for original data. Bottom: Moment 0 results using data with simplification level of 0.0020.

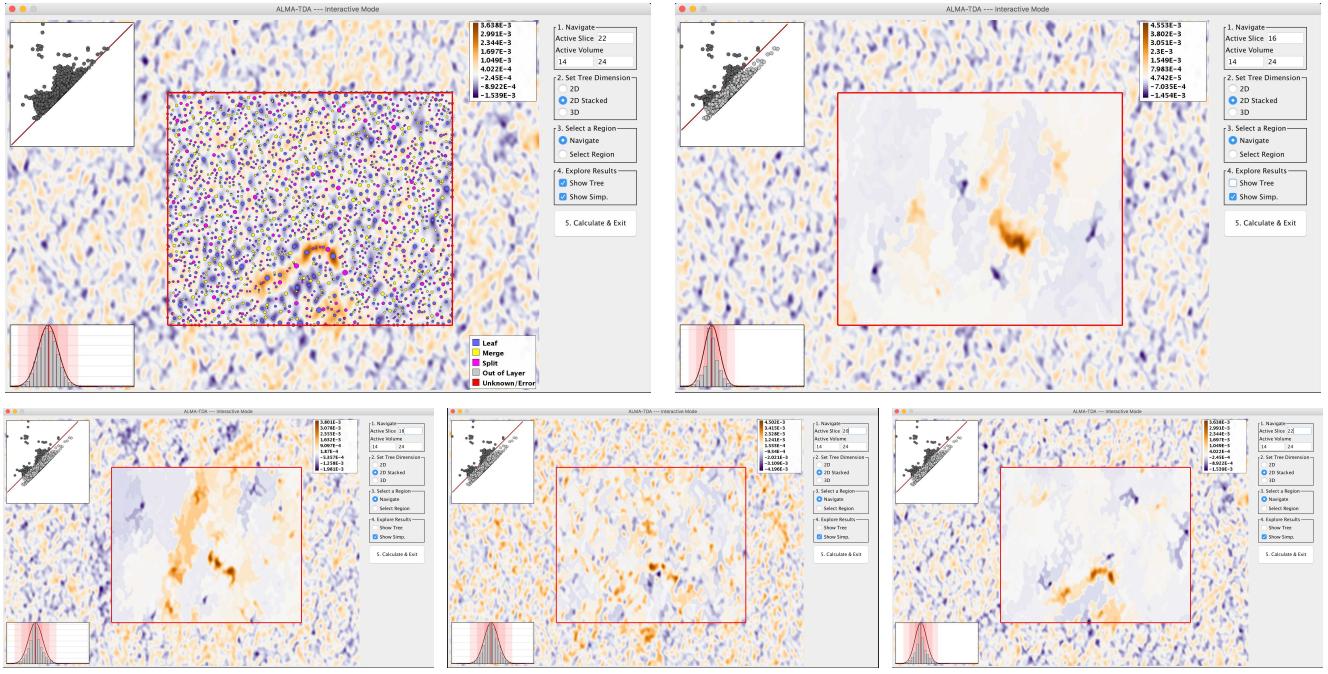


Figure 10: Result of simplifying the Ghost of Mirach dataset using a stack of images with 2D contour trees. Top left: Visualization of the 2D contour tree on slice 22. Top right: Simplification of slice 16. Bottom: Simplification of slices 18, 20, & 22, respectively. The persistent simplification level was 0.00138. The simplification level was good for all except slice 20 where a more aggressive level of simplification was called for.

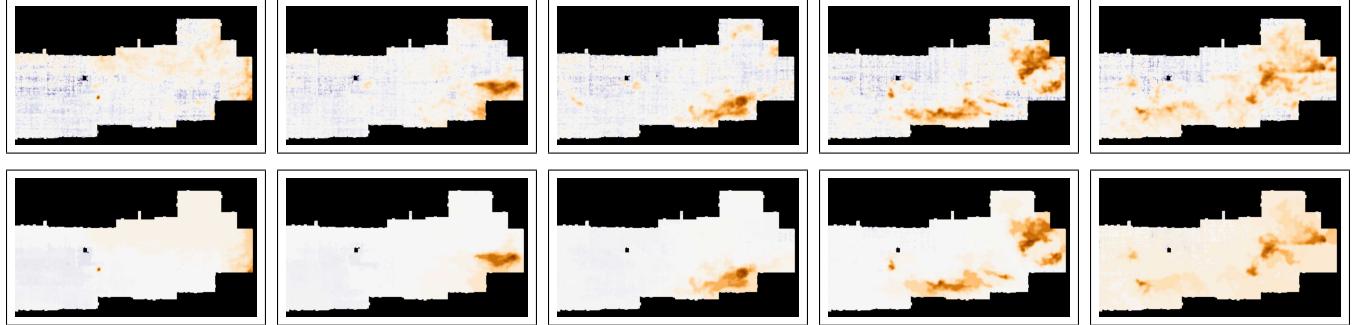


Figure 11: Visualizations of selected slices from the range 100 to 200 of the CMZ data. Top: Slices 100, 120, 140, 160, and 180 before simplification, respectively. Bottom: Slices 100, 120, 140, 160, and 180 after simplification, respectively. The simplification level used was 3.45.

simplification is shared between all slices. This works well for slices 16, 18, and 22 (top right, bottom left, and bottom right, respectively). However, the level of simplification was not aggressive enough for slice 20 (bottom middle). At this point the user could either select a more aggressive simplification, or they could choose to simplify slice 20 separately from the others.

Moment 0 Analysis. Astrophysicists often use what is known as moment analysis to reduce the 3D spectrum to 2D images. Moment 0, 1, and 2 measure the mass of gas, the direction of gas movement, and the temperature of gas, respectively. They are all integrals across the spectra. To demonstrate the noise reducing power of our approach, we show the result of moment 0 analysis in Figure 9 on the 2D stack simplification from Figure 10. Moment 0 is calculated as $m_0 = \int I_v$, where I is the intensity for a given spectra v . By removing the noise from each of the layers, the resulting moment map is significantly less noisy making the signal itself very apparent. Our collaborator also found the dim feature pointed to by the arrow very interesting. He and his collaborators have been actively debating whether this structure is signal or a data processing

artifact. Nevertheless, our approach retained it as signal, and we are excited to see how our results generate further conversations regarding the data.

6.2 CMZ Data Set

The CMZ data are a ^{13}CO 2-1 image of the Central Molecular Zone (CMZ) of the galaxy (data are published in [24]). The data cube is approximate 500MB with resolution of 1150x200 in the spatial domain and 500 in the spectral domain (i.e., 500 slices). We look at 100 slices of a region with resolution about 300x200.

Science Description. The cube shows the low-density molecular gas in the Galaxy's center, with higher intensities generally indicating that there is more gas moving at a particular velocity along each line of sight. It contains highly turbulent gas with properties that are very different than the rest of the Galaxy. The domain scientists use these data to measure the structure of the interstellar medium, which is important for determining how stars are formed and how galaxies evolve. Because the gas they are seeing is in diffuse clouds that do not have well-defined edges, signal identification

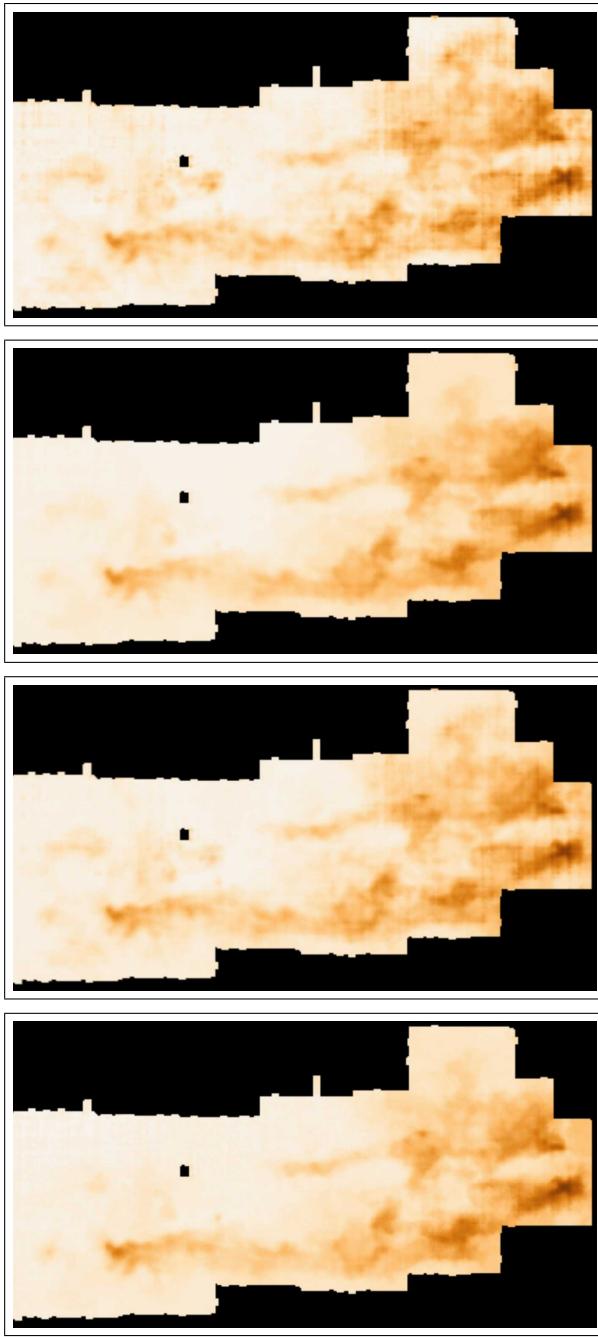


Figure 12: Visualizations of moment 0 for slices 100 to 200 of the CMZ data computed using 2D contour trees. Top: Moment 0 on the original data. Row 2: Moment 0 on all 100 simplified slices. Row 3: Moment 0 only using every 4th slice. Row 4: Moment 0 only using every 8th slice. The simplification level used was 3.45.

is a critical component in improving their understanding of how the gas changes states. Identifying structures in the gas is useful for determining how turbulent it is on different scales, which plays a key role in many star formation theories.

Denoising Slices. Figure 11 shows a number of slices denoised. The signal to noise ratio in this dataset is much better than the previous ones. Nevertheless, many low persistent features have been removed using our approach.

Denoising for Moment Analysis. Deep cubes (those with many slices) such as this one are often created in order to mitigate the impact of noise during moment analysis – by more densely sampling the frequency domain, noise from any single slice has a smaller impact in the output. However, creating deep cubes such as this is computationally and manpower expensive. NRAO has significant human and computational infrastructure dedicated to generating data cubes from the raw data captured by radio telescopes. By providing strong denoising capabilities, data cubes can be sampled at lower spectral frequencies and still produce similar moment maps. See Figure 12 for an example. Here, the top shows the moment map on the original data. Then, moment maps are shown that are calculated using every slice (100 total), every 4th slices (25 total), and every 8th slice (12 total), top to bottom, respectively. The results using fewer slices are virtually indistinguishable from the version using all 100 slices.

7 DISCUSSION

In this feasibility study, we have focused on persistence-based simplification of ALMA data. Our design has focused on the usability objectives of our domain partners, *simplicity, integrability, and reproducibility*, and we recommend these design objectives for anyone else wishing to collaborate with astrophysicists. Thus far, reception of our approach has been good. Virtually everyone who has seen the results are impressed, for some, almost to the point to skepticism. Public outreach with such a new tool using unfamiliar techniques remains challenging. Among astrophysicists, there is a desire to understand both the tool and the underlying technique, and given the complexities of topological data analysis, this can be a challenging, but potentially transformative undertaking.

REFERENCES

- [1] Astrodendro, a python package to compute dendograms of astronomical data. <http://www.dendrograms.org/>.
- [2] QFitsView. <http://www.mpe.mpg.de/~ott/QFitsView/>.
- [3] SAOImage DS9. <http://ds9.si.edu/site/Home.html>.
- [4] P. K. Agarwal, L. Arge, T. Mølhave, M. Revsbæk, and J. Yang. Maintaining contour trees of dynamic terrains. *Proceedings 31st International Symposium on Computational Geometry*, pp. 796–81, 2015.
- [5] C. L. Bajaj, V. Pascucci, and D. R. Schikore. The contour spectrum. *Proceedings 8th conference on Visualization*, pp. 167–ff, 1997.
- [6] A. J. Barth, J. Darling, A. J. Baker, B. D. Boizelle, D. A. Buote, L. C. Ho, and J. L. Walsh. Toward precision black hole masses with alma: Ngc 1332 as a case study in molecular disk dynamics. *The Astrophysical Journal*, 823(1), 2016.
- [7] A. Basmajian and D. Saric. Geodesically complete hyperbolic structures. arxiv.org/abs/1508.02280, 2016.
- [8] A. Belloche, R. T. Garrod, H. S. P. Müller, and K. M. Menten. Detection of a branched alkyl molecule in the interstellar medium: iso-propyl cyanide. *Science*, 345:1584–1587, Sept. 2014. doi: 10.1126/science.1256678
- [9] D. Berry. Fellwalker - a clump identification algorithm. *Astronomy and Computing*, 10:22–31, 2015.
- [10] A. D. Bolatto, S. R. Warren, A. K. Leroy, F. Walter, S. Veilleux, E. C. Ostriker, J. Ott, M. Zwaan, D. B. Fisher, A. Weiss, E. Rosolowsky, and J. Hodge. Suppression of star formation in the galaxy NGC 253 by a starburst-driven molecular wind. *Nature*, 499:450–453, July 2013. doi: 10.1038/nature12351
- [11] H. Carr, C. Sewell, L.-T. Lo, and J. Ahrens. Hybrid Data-Parallel Contour Tree Computation. In C. Turkay and T. R. Wan, eds., *Computer Graphics and Visual Computing (CGVC)*. The Eurographics Association, 2016.
- [12] H. Carr and J. Snoeyink. Path seeds and flexible isosurfaces using topology for exploratory visualization. *Proceedings of the Symposium on Data Visualisation*, 2003.
- [13] H. Carr, J. Snoeyink, and U. Axen. Computing contour trees in all dimensions. *Computational Geometry: Theory and Applications*, 24(3):75–94, 2003.

- [14] H. Carr, J. Snoeyink, and M. van de Panne. Simplifying flexible isosurfaces using local geometric measures. *Proceedings 15th IEEE Visualization*, pp. 497–504, 2004.
- [15] H. Carr, J. Snoeyink, and M. van de Panne. Flexible isosurfaces: Simplifying and displaying scalar topology using the contour tree. *Computational Geometry: Theory and Applications*, 43(1):42–58, 2010.
- [16] H. Carr, G. Weber, C. Sewell, and J. Ahrens. Parallel peak pruning for scalable smp contour tree computation. In *IEEE Symposium on Large Data Analysis and Visualization*, 2016.
- [17] Y.-J. Chiang, T. Lenz, X. Lu, and G. Rote. Simple and optimal output-sensitive construction of contour trees using monotone paths. *Computational Geometry*, 30(2):165 – 195, 2005.
- [18] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete and Computational Geometry*, 37(1):103–120, 2007.
- [19] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Extending persistence using poincaré and lefschetz duality. *Foundations of Computational Mathematics*, 9(1):79–103, 2009.
- [20] D. Colombo, E. Rosolowsky, A. Ginsburg, A. Duarte-Cabral, and A. Hughes. Graph-based interpretation of the molecular interstellar medium segmentation. *Monthly Notices of the Royal Astronomical Society*, 454(2):2067–2091, 2015.
- [21] C. De Breuck, R. J. Williams, M. Swinbank, P. Caselli, K. Coppin, T. A. Davis, R. Maiolino, T. Nagao, I. Smail, F. Walter, A. Weiss, and M. A. Zwaan. ALMA resolves turbulent, rotating [CII] emission in a young starburst galaxy at $z = 4.8$. *Astronomy & Astrophysics*, 565:A59, May 2014. doi: 10.1051/0004-6361/201323331
- [22] H. Edelsbrunner, D. Letscher, and A. J. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28:511–533, 2002.
- [23] H. Edelsbrunner, D. Morozov, and V. Pascucci. Persistence-sensitive simplification of functions on 2-manifolds. *Proceedings of the Annual ACM Symposium on Computational Geometry*, pp. 127–134, 2006.
- [24] A. Ginsburg, C. Henkel, Y. Ao, D. Riquelme, J. Kauffmann, T. Pillai, E. A. Mills, M. A. Requena-Torres, K. Immer, L. Testi, et al. Dense gas in the galactic central molecular zone is warm and heated by turbulence. *Astronomy & Astrophysics*, 586:A50, 2016.
- [25] A. A. Goodman, E. W. Rosolowsky, M. A. Borkin, J. B. Foster, M. Halle, J. Kauffmann, and J. E. Pineda. A role for self-gravity at multiple length scales in the process of star formation. *Nature*, 457(7225):63–66, 2009.
- [26] C. Gueunet, P. Fortin, J. Jomier, and J. Tierny. Contour forests: Fast multi-threaded augmented contour trees. In *IEEE Symposium on Large Data Analysis and Visualization*, 2016.
- [27] P. J. Hancock, T. Murphy, B. M. Gaensler, A. Hopkins, and J. R. Curran. Compact continuum source-finding for next generation radio surveys. *Monthly Notices of the Royal Astronomical Society*, 422(2), 2012.
- [28] A. M. Hopkins, M. T. Whiting, N. Seymour, K. E. Chow, R. P. Norris, L. Bonavera, R. Breton, D. Carbone, C. Ferrari, T. M. O. Franzen, H. Garsden, J. González-Nuevo, C. A. Hales, P. J. Hancock, G. Heald, D. Herranz, M. Huynh, R. J. Jurek, M. López-Caniego, M. Massardi, N. Mohan, S. Molinari, E. Orrù, R. Paladino, M. Pestalozzi, R. Pizzo, D. Rafferty, H. J. A. Röttgering, L. Rudnick, A. Schisano, E. Shulevski, J. Swinbank, R. Taylor, and A. J. van der Horst. The askap/emu source finding data challenge. *Publications of the Astronomical Society of Australia*, 32, 2015.
- [29] K. E. Johnson, A. K. Leroy, R. Indebetouw, C. L. Brogan, B. C. Whitmore, J. Hibbard, K. Sheth, and A. S. Evans. The Physical Conditions in a Pre-super Star Cluster Molecular Cloud in the Antennae Galaxies. *The Astrophysical Journal*, 806:35, June 2015. doi: 10.1088/0004-637X/806/1/35
- [30] A. K. Leroy, A. D. Bolatto, E. C. Ostriker, E. Rosolowsky, F. Walter, S. R. Warren, J. Donovan Meyer, J. Hodge, D. S. Meier, J. Ott, K. Sandstrom, A. Schruba, S. Veilleux, and M. Zwaan. ALMA Reveals the Molecular Medium Fueling the Nearest Nuclear Starburst. *The Astrophysical Journal*, 801:25, Mar. 2015. doi: 10.1088/0004-637X/801/1/25
- [31] H. B. Liu, R. Galván-Madrid, I. Jiménez-Serra, C. Román-Zúñiga, Q. Zhang, Z. Li, and H.-R. Chen. ALMA Resolves the Spiraling Accretion Flow in the Luminous OB Cluster-forming Region G33.92+0.11. *The Astrophysical Journal*, 804:37, May 2015. doi: 10.1088/0004-637X/804/1/37
- [32] S. Maadasamy, H. Doraiswamy, and V. Natarajan. A hybrid parallel algorithm for computing and tracking level set topology. In *HiPC*, pp. 1–10. IEEE Computer Society, 2012.
- [33] S. McKenna, D. Mazur, J. Agutter, and M. Meyer. Design activity framework for visualization design. *IEEE Transactions on Visualization & Computer Graphics*, 20:2191–2200, 2014.
- [34] D. S. Meier, F. Walter, A. D. Bolatto, A. K. Leroy, J. Ott, E. Rosolowsky, S. Veilleux, S. R. Warren, A. Weiss, M. A. Zwaan, and L. K. Zschaechner. ALMA Multi-line Imaging of the Nearby Starburst NGC 253. *The Astrophysical Journal*, 801:63, Mar. 2015. doi: 10.1088/0004-637X/801/1/63
- [35] E. Merényi, J. Taylor, and A. Isella. Mining complex hyperspectral alma cubes for structure with neural machine learning. *IEEE Symposium Series on Computational Intelligence (SSCI)*, 2016.
- [36] D. Morozov and G. Weber. Distributed merge trees. In *Proceedings of the 18th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*, pp. 93–102. New York, NY, USA, 2013.
- [37] D. Morozov and G. H. Weber. *Distributed Contour Trees*, pp. 89–102. Springer International Publishing, Cham, 2014.
- [38] D. D. Nguyen, A. C. Seth, M. den Brok, N. Neumayer, M. Cappellari, A. J. Barth, N. Caldwell, B. F. Williams, and B. Binder. Improved dynamical constraints on the mass of the central black hole in ngc 404. *The Astrophysical Journal*, 836(2), 2017.
- [39] P. Oesterling, C. Heine, G. Weber, D. Morozov, and G. Scheuermann. Computing and visualizing time-varying merge trees for high-dimensional data. Workshop on Topology-Based Methods in Visualization (TopoInVis), May 2015.
- [40] K. Onishi, S. Iguchi, T. A. Davis, M. Bureau, M. Cappellari, M. Sarzi, and L. Blitz. Wisdom project - i: Black hole mass measurement using molecular gas kinematics in ngc 3665. *Accepted to Monthly Notices of the Royal Astronomical Society (MNRAS)*, 2017.
- [41] A. Partnership, C. L. Brogan, L. M. Perez, T. R. Hunter, W. R. F. Dent, A. S. Hales, R. Hills, S. Corder, E. B. Fomalont, C. Vlahakis, Y. Asaki, D. Barkats, A. Hirota, J. A. Hodge, C. M. V. Impellizzeri, R. Kneissl, E. Liuzzo, R. Lucas, N. Marcelino, S. Matsushita, K. Nakaniishi, N. Phillips, A. M. S. Richards, I. Toledo, R. Aladro, D. Broguiere, J. R. Cortes, P. C. Cortes, D. Espada, F. Galarza, D. Garcia-Appadoo, L. Guzman-Ramirez, E. M. Humphreys, T. Jung, S. Kameno, R. A. Laing, S. Leon, G. Marconi, A. Mignano, B. Nikolic, L.-A. Nyman, M. Radiszc, A. Remijan, J. A. Rodon, T. Sawada, S. Takahashi, R. P. J. Tilanus, B. Vila Vilaro, L. C. Watson, T. Wiklind, E. Akiyama, E. Chapillon, I. de Gregorio-Monsalvo, J. Di Francesco, F. Gueth, A. Kawamura, C.-F. Lee, Q. Nguyen Luong, J. Mangum, V. Pietu, P. Sanhueza, K. Saigo, S. Takakuwa, C. Ubach, T. van Kempen, A. Wootten, A. Castro-Carrizo, H. Francke, J. Gallardo, J. Garcia, S. Gonzalez, T. Hill, T. Kaminski, Y. Kurono, H.-Y. Liu, C. Lopez, F. Morales, K. Plarre, G. Schieven, L. Testi, L. Videla, E. Villard, P. Andreani, J. E. Hibbard, and K. Tatematsu. First Results from High Angular Resolution ALMA Observations Toward the HL Tau Region. *ArXiv e-prints*, Mar. 2015.
- [42] A. Partnership, C. Vlahakis, T. R. Hunter, J. A. Hodge, L. M. Pérez, P. Andreani, C. L. Brogan, P. Cox, S. Martin, M. Zwaan, S. Matsushita, W. R. F. Dent, C. M. V. Impellizzeri, E. B. Fomalont, Y. Asaki, D. Barkats, R. E. Hills, A. Hirota, R. Kneissl, E. Liuzzo, R. Lucas, N. Marcelino, K. Nakaniishi, N. Phillips, A. M. S. Richards, I. Toledo, R. Aladro, D. Broguiere, J. R. Cortes, P. C. Cortes, D. Espada, F. Galarza, D. Garcia-Appadoo, L. Guzman-Ramirez, A. S. Hales, E. M. Humphreys, T. Jung, S. Kameno, R. A. Laing, S. Leon, G. Marconi, A. Mignano, B. Nikolic, L.-A. Nyman, M. Radiszc, A. Remijan, J. A. Rodón, T. Sawada, S. Takahashi, R. P. J. Tilanus, B. Vila Vilaro, L. C. Watson, T. Wiklind, Y. Ao, J. Di Francesco, B. Hatsukade, E. Hatziminaoglou, J. Mangum, Y. Matsuda, E. van Kampen, A. Wootten, I. de Gregorio-Monsalvo, G. Dumas, H. Francke, J. Gallardo, J. Garcia, S. Gonzalez, T. Hill, D. Iono, T. Kaminski, A. Karim, M. Krips, Y. Kurono, C. Lonsdale, C. Lopez, F. Morales, K. Plarre, L. Videla, E. Villard, J. E. Hibbard, and K. Tatematsu. ALMA Long Baseline Observations of the Strongly Lensed Submillimeter Galaxy HATLAS J090311.6+003906 at $z=3.042$. *ArXiv e-prints*, Mar. 2015.

- [43] V. Pascucci, K. Cole-McLaughlin, and G. Scorzelli. Multi-resolution computation and presentation of contour trees. Technical Report UCRL-PROC-208680, Lawrence Livermore National Laboratory, 2005.
- [44] V. Pascucci, G. Scorzelli, P.-T. Bremer, and A. Mascarenhas. Robust on-line computation of Reeb graphs: Simplicity and speed. *ACM Transactions on Graphics*, 26(3):58.1–58.9, 2007.
- [45] N. Peretto, G. A. Fuller, A. Duarte-Cabral, A. Avison, P. Hennebelle, J. E. Pineda, P. André, S. Bontemps, F. Motte, N. Schneider, and S. Molinari. Global collapse of molecular clouds as a formation mechanism for the most massive stars. *Astronomy & Astrophysics*, 555:A112, July 2013. doi: 10.1051/0004-6361/201321318
- [46] B. Raichel and C. Seshadhri. A mountaintop view requires minimal sorting: A faster contour tree algorithm. *CoRR*, abs/1411.2689, 2014.
- [47] J. M. Rathborne, S. N. Longmore, J. M. Jackson, J. F. Alves, J. Bally, N. Bastian, Y. Contreras, J. B. Foster, G. Garay, J. M. D. Kruijssen, L. Testi, and A. J. Walsh. A Cluster in the Making: ALMA Reveals the Initial Conditions for High-mass Cluster Formation. *The Astrophysical Journal*, 802:125, Apr. 2015. doi: 10.1088/0004-637X/802/2/125
- [48] J. M. Rathborne, S. N. Longmore, J. M. Jackson, J. M. D. Kruijssen, J. F. Alves, J. Bally, N. Bastian, Y. Contreras, J. B. Foster, G. Garay, L. Testi, and A. J. Walsh. Turbulence Sets the Initial Conditions for Star Formation in High-pressure Environments. *The Astrophysical Journal Letters*, 795:L25, Nov. 2014. doi: 10.1088/2041-8205/795/2/L25
- [49] P. Rosen, J. Tu, and L. Piegl. A hybrid solution to calculating augmented join trees of 2d scalar fields in parallel. In *CAD Conference and Exhibition (accepted)*, 2017.
- [50] E. Rosolowsky and A. Leroy. Bias-free measurement of giant molecular cloud properties. *The Publications of the Astronomical Society of the Pacific*, 118(842):590–610, 2006.
- [51] E. W. Rosolowsky, J. E. Pineda, J. Kauffmann, and A. A. Goodman. Structural analysis of molecular clouds: Dendograms. *The Astrophysical Journal*, 679(2):1338–1351, 2008.
- [52] D. Schneider, A. Wiebel, H. Carr, M. Hlawitschka, and G. Scheuermann. Interactive comparison of scalar fields based on largest contours with applications to flow visualization. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1475–1482, 2008.
- [53] B.-S. Sohn and C. Bajaj. Time-varying contour topology. *IEEE Transactions on Visualization and Computer Graphics*, 12(1):14–25, 2006.
- [54] M. van Kreveld, R. van Oostrum, C. L. Bajaj, V. Pascucci, and D. Schikore. Contour trees and small seed sets for isosurface traversal. *Proceedings 19th Annual symposium on Computational geometry*, pp. 212–220, 1997.
- [55] R. Wang, J. Wagg, C. L. Carilli, F. Walter, L. Lentati, X. Fan, D. A. Riechers, F. Bertoldi, D. Narayanan, M. A. Strauss, P. Cox, A. Omont, K. M. Menten, K. K. Knudsen, R. Neri, and L. Jiang. Star Formation and Gas Kinematics of Quasar Host Galaxies at $z \sim 6$: New Insights from ALMA. *The Astrophysical Journal*, 773:44, Aug. 2013. doi: 10.1088/0004-637X/773/1/44
- [56] S. Westerlund, C. Harris, and T. Westmeier. Assessing the accuracy of radio astronomy source finding algorithms. *Publications of the Astronomical Society of Australia*, 29:301–308, 2012.
- [57] J. P. Williams, E. J. de Geus, and L. Blitz. Determining structure in molecular clouds. *The Astrophysical Journal*, 428(2):693–712, 1994.