**NASA DEVELOP National Program**



NASA Langley Research Center

*Summer 2017*

CALIPSO Cross-Cutting

Enhancements to Visualization of CALIPSO (VOCAL) through Case Studies of Saharan Dust

 **Technical Report**

Final Draft – August 9, 2017

Collin Pampalone (Project Lead)

Ryan Avery

William Turner

Dr. Kathleen Moore, NASA Langley Research Center (Science Advisor)

Dr. Kenton Ross, NASA Langley Research Center (Science Advisor)

Previous Contributors:

Grant Mercer

Nathan Qian

Ashna Aggarwal

Jordan Vaa

Kathleen Moore

Ahmad Aburizaiza

Andrea Martinez

Adama Ba

Joseph Driscoll

# 1. Abstract

The Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite’s CALIOP sensor generates vertical Lidar profiles of the atmosphere at a global scale. Currently, the standard visualization tool for these data is written in Interactive Data Language (IDL), a proprietary language that does not support features for tracking aerosols, selecting data, or sharing those selected sections. This makes working with CALIPSO data difficult for researchers and does not allow them to visually identify aerosol features from these data. Previous DEVELOP teams have built a working version of the *Visualization of CALIPSO* (VOCAL) software, a Python language replacement for this IDL-based software. During this term, the team enhanced VOCAL by improving the shape drawing tool, adding the capability to view multiple levels of data, and more flexible data inputs and outputs that support a decentralized database. These features will increase the usability of VOCAL, expediting the process of visually identifying features and analyzing the resulting subsets of data. Finally, the DEVELOP team conducted a case study with Saharan dust transport over the Atlantic Ocean to test the capabilities of the VOCAL software and to produce a database of dust events. The database and case study will help members of the CALIPSO Science Team compare the performance of classification algorithms used to create Level 2 (L2) CALIPSO data products, and will also support preliminary analysis of the atmospheric interactions and consequences related to long range dust transport.

**Keywords**

Aerosol, VOCAL, CALIPSO, CALIOP, Lidar, Saharan dust, open-source, database

# 2. Introduction

* 1. ***Background Information***

Aerosols, or small particles suspended in the air, are an important component influencing Earth’s atmosphere. They come from many different sources, both natural and artificial, and can affect everything from the earth’s radiation budget, to cloud formation, to nutrient transport (Fakowski, 1998). Commonly known anthropogenic aerosols include sulfate from the burning of fossil fuels, chlorofluorocarbons from consumer products like hairspray and refrigerants, and soot from biomass burning (Voiland, 2010). A less well-known, non-anthropogenic aerosol is atmospheric dust, which in the summer months is transported in great densities over large distances across the Atlantic Ocean (Allen, 2015). The Sahara Desert is the largest source of atmospheric dust, and previous studies have examined the impact this dust has on regulating the growth of primary producers in the Western Pacific Ocean (Falkowski et al., 1998; Liu et al., 2008). A better understanding of the optical properties of dust is required to identify and characterize dust features, in order to increase the body of knowledge on the effects that desert dust has on the climate and biosphere. To this end, the CALIPSO Cross-Cutting team designed three goals: (1) improve the *Visualization of CALIPSO* (VOCAL) software, an interactive GUI used to display CALIPSO curtain data and manually digitize atmospheric features, (2) create a MySQL database of dust features using our improved tool, and (3) conduct a case study of ten long range dust transport events that occurred between 2014 and 2016 to determine the difference in trends between visually identified features and CALIPSO’s Level 2, 5 km Merged Layer product.

The Cloud Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) mission was launched in 2006, with its primary sensor the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP). While other sensors often gather imagery data in the horizontal plane, CALIOP records sets of vertical profiles, or “curtains”, of backscatter and depolarization measurements across the globe (Winker et al., 2006, 2007). These vertical curtains give researchers unique insight into the vertical composition of the atmosphere, from altitudes at sea-level to 30 km high. CALIOP uses 532 nm and 1064 nm lasers to penetrate through aerosols and clouds, and records the time lapsed, backscatter amount, and depolarization of the backscattered light. These measurements give researchers the ability to view inside or even through aerosol layers, as well as to measure optical properties that can be used to estimate energy absorption and reflectance (Winker et al., 2010). Level 2 processing of CALIPSO data by algorithms allows researchers to classify aerosol layers by aerosol type, subtype, cloud type, and ice/water phase (Stuart et al., 2004, Omar et al., 2009). However, these algorithms continue to be improved upon, and have documented issues with correctly parsing multiple features and correctly classifying certain aerosol types (Vaughan et al., 2004). To inform future development, it is necessary to compare the performance of these algorithms to some baseline in order to investigate where bias and error is introduced. Our study provides such a baseline for comparison in our database of visually identified features, which were created using VOCAL.

Researchers currently use a set of scripts written in the proprietary language Interactive Data Language (IDL), which do not offer a way to select data. Over several terms DEVELOP teams have created VOCAL as a means to visualize, select, and share CALIPSO data. During the previous term, several features were added to improve user friendliness such as rendering multiple levels of data and an improved drawing tool. However, these tools were never fully implemented and must be completed and debugged. While the tool has already been used to study Level 1 data, more improvements are needed to make the tool more user friendly and to allow the import and interactivity with Level 2 data products. By studying and recording dust events transporting from the west coast of Africa to Central and South America from 2014 to 2016, we produced a database of events and analysis of statistics related to these events that will allow the end user to discern differences between user-identified and algorithmically-identified features. Additionally, this case study and its extensive use of VOCAL will allow us to identify ways to continue to improve the tool as we collect user feedback.

* 1. ***Project Partners & Objectives***

Our project partner, Dr. Charles “Chip” Trepte of the CALIPSO Science Team has asked that we make improvements to the tool so that it can eventually replace IDL as the team’s main visualization tool. Other objectives included: (1) allowing users to display and toggle between Level 2 data products and the ability to draw features that are preserved and displayed when tabs are toggled, (2) implementing a free-hand draw tool to allow for a more precise delineation of user-defined features than our current polygon tool allows, and (3) conducting a case study of long range transport of Saharan dust that will make use of these new features, with the desired output being a database of features and some results and analysis of the differences in statistics between features identified from Level 1 (by the user) and Level 2 (by the algorithms) data.

# 3. Methodology

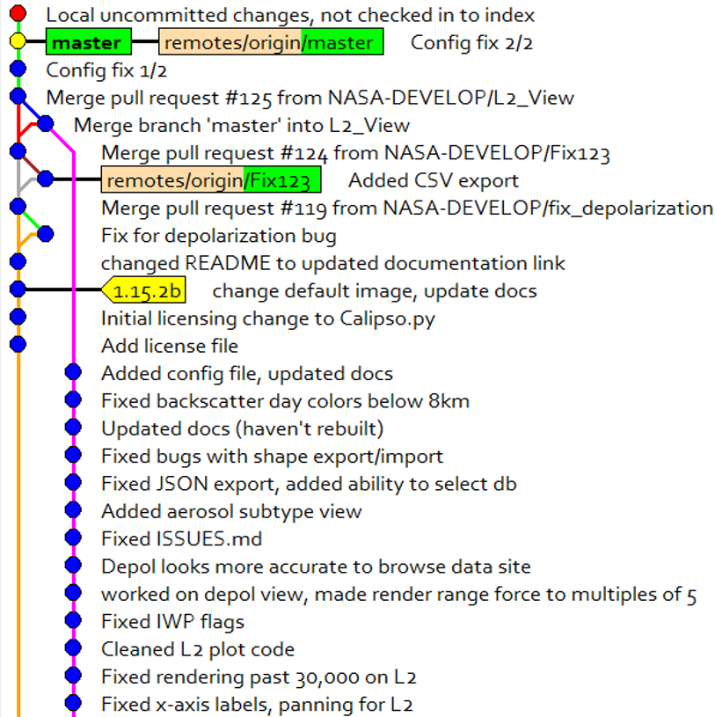
***3.1 VOCAL Development***

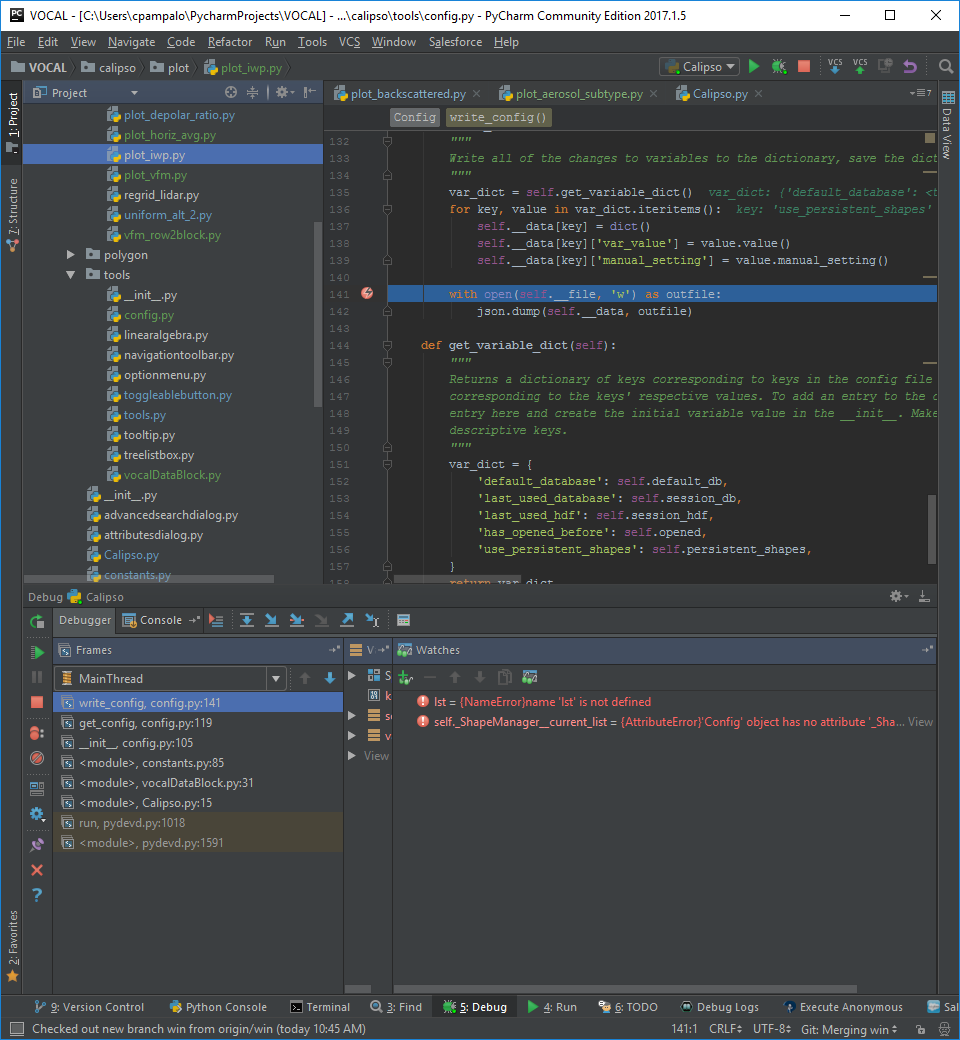
During the previous DEVELOP term, members of the CALIPSO team coded several new features but did not integrate or test them. At the beginning of this term, their files were handed over as a zipped archive. However, previous teams used GitHub, an open source development tool that allows programmers to share source code and do version control online. Additionally, GitHub allows developers to share projects with the public. The first step this term was therefore to fix bugs and integrate features from the previous term into the *master* branch on the GitHub.

We determined which features would be necessary using both our partners’ requests, observed areas of improvement from the case study, and time constraints from the term. Considering these two sources, the features determined to be most necessary were tool integration for all levels, persistent shapes, more dynamic import and export, a settings dialog and configuration file, an improved drawing tool, and updated documentation.

***3.1.1 Version Control***

One of the main issues we encountered in the transition from the last CALIPSO term was issues with the versioning of VOCAL. While they put effort into creating many features, they were not added together in a single usable branch on the GitHub. Instead, their code was scattered in different files and branches. Therefore, we created guidelines for making updates to our own code using Git.





*Figure 1. Left:* The Git history of VOCAL. Branches are shown as lines in different colors and commits as nodes. *Right:* the IDE Pycharm features a debugging tool for Python that helps with development.

In order to make updates to features while maintaining a stable *master* branch (or main, default branch) on the GitHub, we used a set of guidelines to ensure updates happened smoothly. First, the *master* branch would be cloned into a separate branch with an appropriate name for the feature that would be worked on for that branch (e.g. *Persistent\_Shapes*). Then, features would be developed on the new branch, and it would be tested thoroughly. Then, when it was determined that the new feature worked as intended and created no new issues, it was merged into the *master* branch on GitHub. The documentation outlines these coding guidelines and references recources that will be helpful for learning to use Git and Gihub.

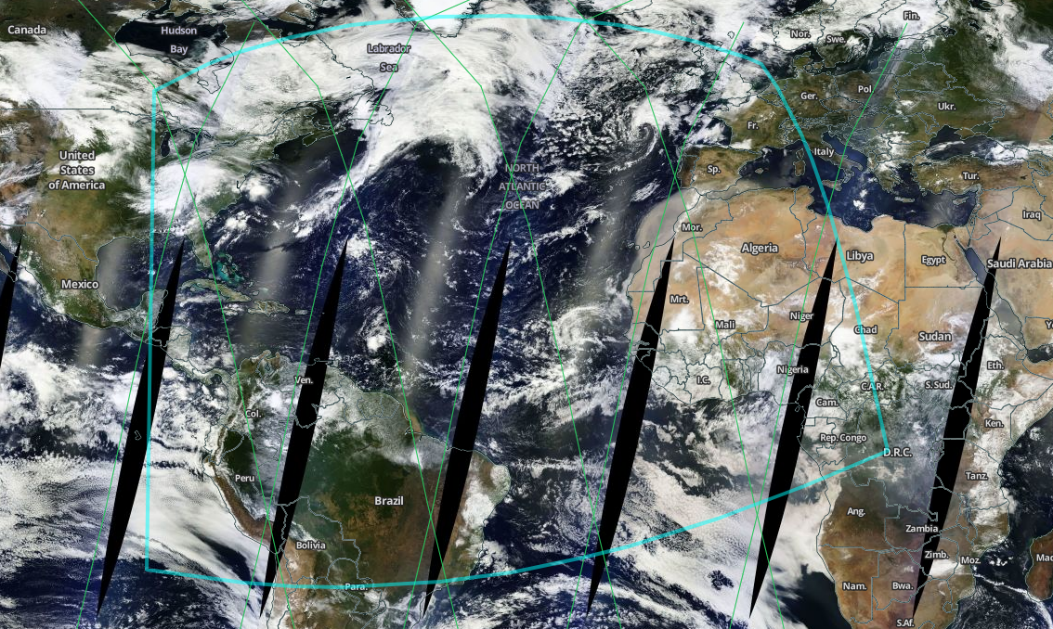
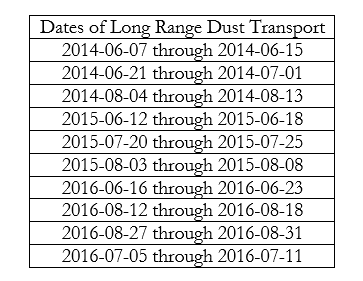
***3.1.2 Debugging***

While most features were added prior to beginning the case study, several were added as a result of areas of improvement discovered while working with VOCAL. Additionally, many bugs became apparent while conducting the case study and could be fixed on the go.

When fixing bugs found in the *master* branch, the typical method of creating a new branch was not followed. Instead, fixes were created, tested, and added directly to the *master* branch without cloning and merging.

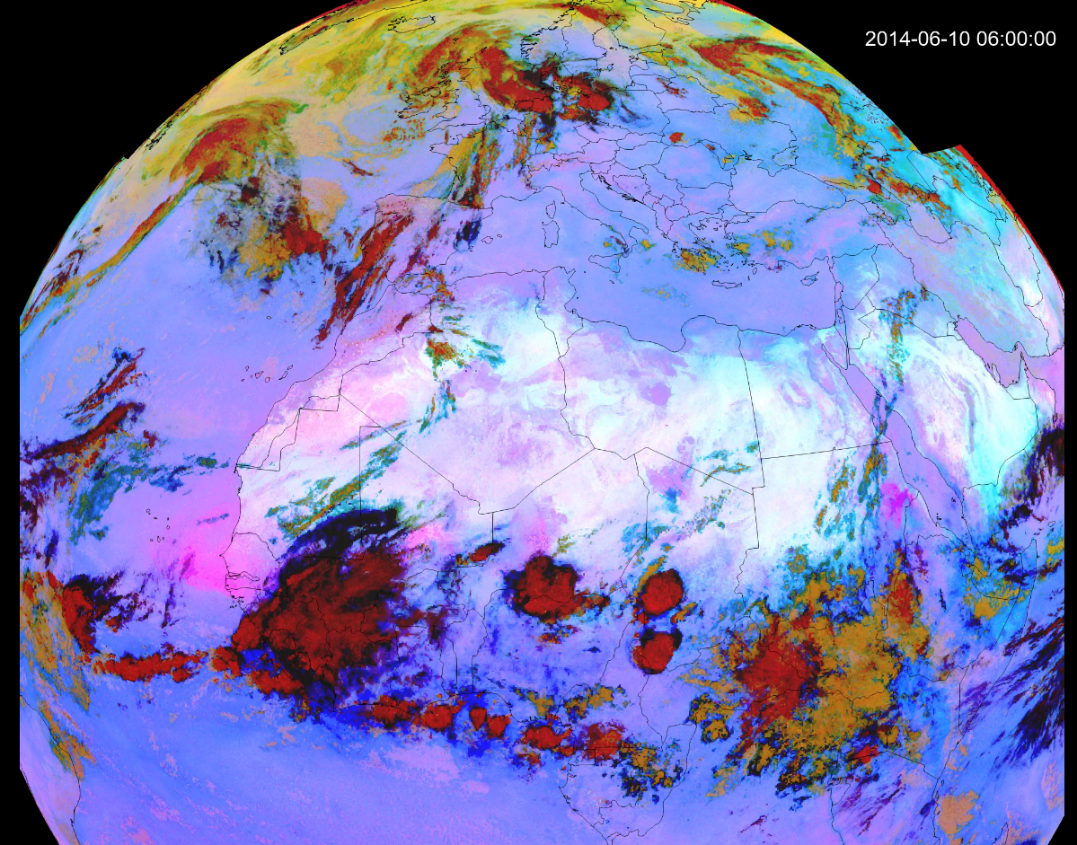
***3.2 Data Acquisition***

The study area encompassed the total area where long range dust transport occurs between the continents of Africa and the Americas (Figure 2). We identified 10 long range dust transport events originating on or near the northwestern coast of Africa during the summer months of 2014, 2015, and 2016 (Table 1). Dust Events were identified using false color visualizations produced by CALIPSO Science Team (Figure 3).

After identifying the date ranges for these dust events, we used NASA’s Earthdata Search tool to find the specific granules of CALIPSO data that intersected each dust event over the course of its transport. We then downloaded these data by requesting these granules using the ASDC HTML Order Tool. The products we acquired is the CAL\_LID\_L1-Standard-V4-10, CAL\_LID\_L2\_05kmMLay-Standard-V4, and CAL\_LID\_L2\_VFM.

*Table 1.* Dates of Long Range Dust Transport Determined from False Color Videos

*Figure 2.* The study area covering the atmospheric corridor where dust transport occurs over the Atlantic. Our bounding polygon is denoted in light blue, CALIPSO granules are represented in green (NASA EarthData).

***3.4 Data Processing***

*Figure 3.* Snapshot offalse color video of atmospheric activity over Africa and the Atlantic. Dust is represented in pink and magenta hues, with the magenta cloud off of the coast in this image showing a severe dust storm. Courtesy of Roman Kowch.

In order to collect data, we first had to identify dust shapes in the CALIPSO curtain data. Using the updated version of VOCAL, we selected shapes we believed to be dust. We based our decision on parameters from our partners and from observations we made using sample data (data that was used for practice and did not affect the results). Our parameters for dust shapes were: medium to high backscatter attenuation between 0.002 and 0.006 km-1sr-1 (displayed as yellow to red in VOCAL), uniform concentration of mostly homogenous backscatter, shapes between 0 and 8km, and shapes between 15°S to 50°N. In contrast to cumulous clouds, which could be found in the same region with similar parameters, dust shapes are much longer, have a smoother surface, and tend to not totally block the Lidar signal. In contrast to smoke, which could often be found slightly south of Saharan dust, dust tends to accumulate lower and spread out across a range of altitudes.

***3.5 Data Analysis***

Once we identified dust shapes across the 10 events, we ran them through a statistical analysis script provided by our partners. Using this script, we iterated through the shapes and pulled data from their corresponding granule files in the region bounded by the shape. Then we compared the data from Level 1 products (total attenuated backscatter and depolarization) to the data in Level 2 products (5km merged layer and vertical feature mask).

# 4. Results & Discussion

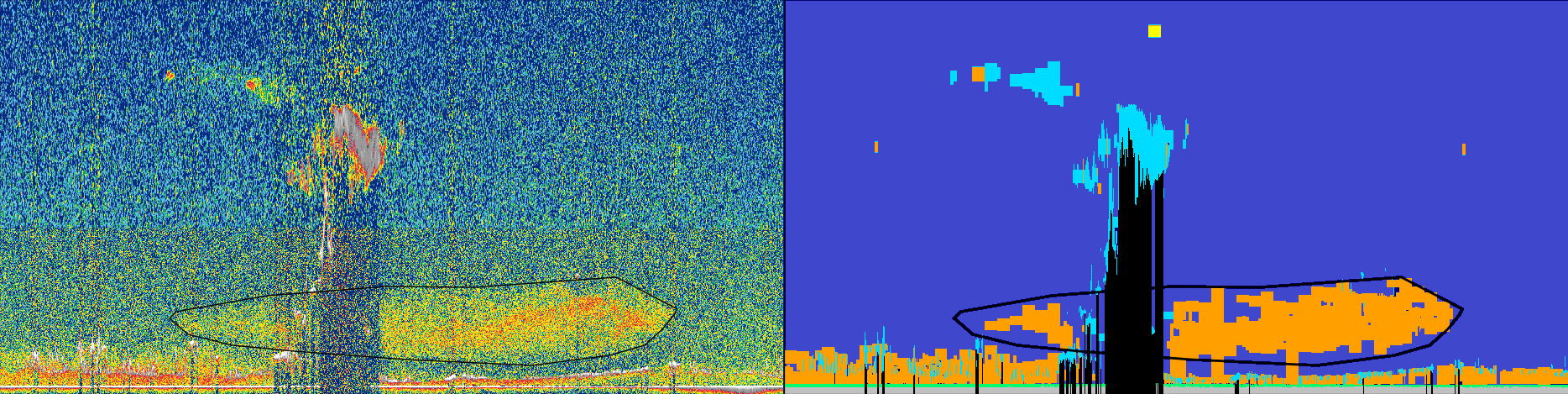
***4.1 VOCAL 1.17.8***

From the previous term, the CALIPSO Cross-Cutting team integrated Level 1 and 2 loading and plotting. Prior to integration, VOCAL could only display the total backscatter attenuation and depolarization plots from a Level 1 CALIPSO data file. Now, VOCAL allows users to load Level 1 and Level 2 Data files and view data plots from each including backscatter, depolarization, vertical feature mask, horizontal averaging, ice/water phase, and aerosol subtype.

VOCAL also includes features newly developed during the Summer 2017 term including tool integration for all levels of plots, persistent shapes, more dynamic import and export, a settings dialog and configuration file, an improved drawing tool, and updated documentation including instructions for both end users and future developers.

The new tool integration and persistent shapes feature comes as part of an effort to make VOCAL more intuitive and faster to use for researchers. From recommendations from end users and experience with VOCAL during the case study, the CALIPSO Cross-Cutting team determined that tool integration that traversed levels would be important in making the tool more efficient. Now, when tools interact with one level of VOCAL, for example, to pan or create a new shape, those changes appear when the plot is switched to a new view. These make it easier to switch between plots to view differences or similarities, or to verify selections using Level 2 data.

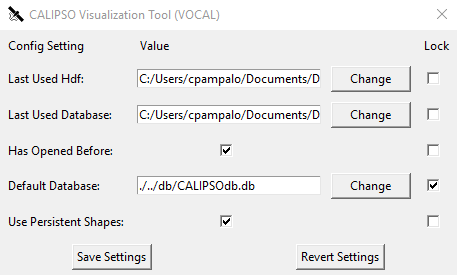
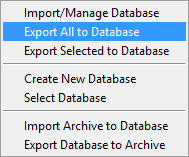
*Figure 4:* *Left:* A backscatter plot showing dust selected with the free draw tool. *Right:* the same shape and special region shown when switched to vertical feature mask. Note that the orange selected in VFM corresponds to tropospheric aerosol. However, in the backscatter plot, only higher attenuation is shown.



New import and export features allow users to interact with CALIPSO polygon databases in several new ways (Figure 5, right) Users can now create new databases in a user-determined folder and switch between databases. Therefore, researchers can more easily access their database to extract polygon data.

The new settings dialog (Figure 5, left) and configuration file gives VOCAL users access to some of the settings within VOCAL. Settings include the default database, which is the database that VOCAL will save and open polygons from when started, a persistent shapes checkbox, which allows users to switch between persistent shapes or new shapes for each view, and several default folders, which allow the user to specify where VOCAL should open the file browser.

The new drawing tool makes the old tool more similar to the polygon tool found in image processing software. Users click to select a vertex and a bounding line appears where the line will be drawn.

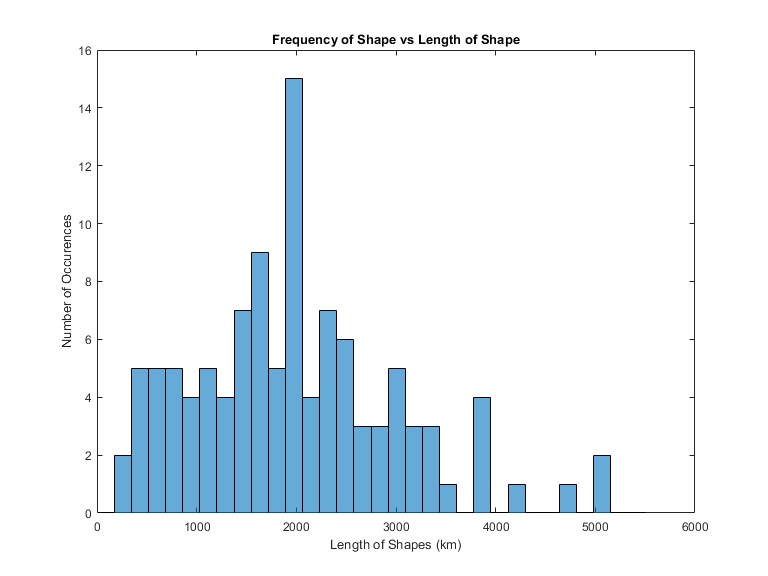
*Figure 5:* *Left:* The settings dialog displaying a list of options. Locking the value forces to program to only allow manual changes to it. Some values, such as *Last Used Database* are changed by VOCAL. *Right:* the database menu with added *Create New Database* and *Select Database*.

Finally, updated documentation helps both end users and developers access the software. Prior to the current term, the documentation had not been updated since 2015. Many updates have been made since then in both the interface and structure of the program. Therefore, the new documentation includes guides for accessing the latest version of VOCAL from GitHub, installing VOCAL, reporting issues, making updates to VOCAL, and properly managing and releasing versions.

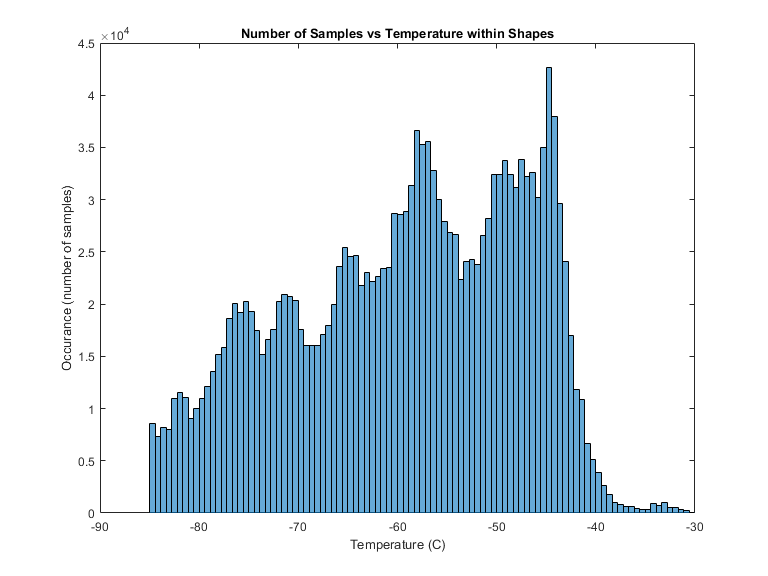
***4.2 Analysis of Results***

We produced several graphs comparing shape and layer data to explain our results of the case study. Shapes are the polygons we draw in VOCAL around dust data, while layers are average objects identified by the algorithms and published in Level 2 data. While the larger implication of these graphs and the data will be determined by our partner, we can explain what the graphs show.

We first graphed the frequency of shapes vs. the length of those shapes (Figure 6). Shape length spiked around 2000km but was mostly spread from 0 to 3500km, with no shapes larger than 5500km. Next, we graphed the frequency of temperature by sample vs. temperature of the sample (see Figure 7). This histogram shows the distribution of temperatures within the drawn shapes. As shown, there is a spike near 45°C which slowly decreases as temperature goes down.

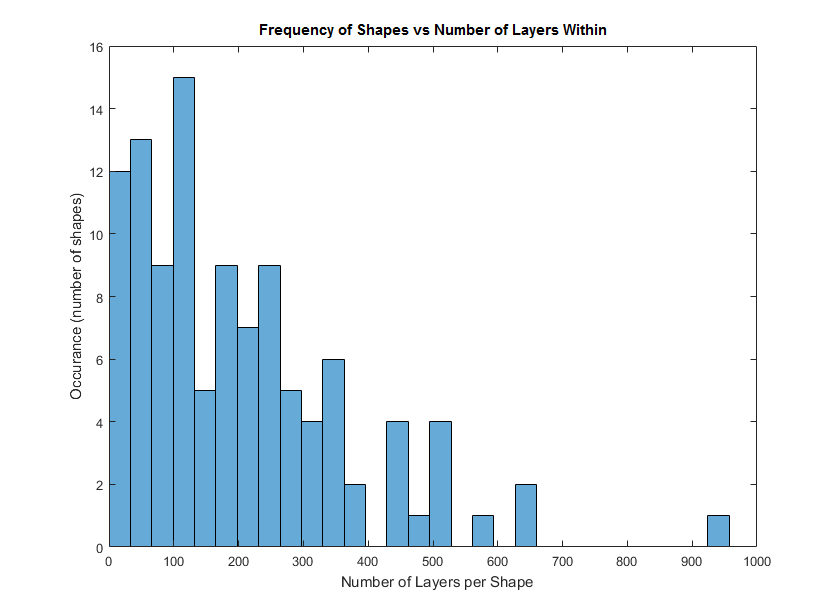


*Figure 6:* Frequency of shapes vs the length of the shape. This histogram shows the distribution of lengths from the drawn shapes across all of the events.



*Figure 7:* Frequency of sample vs the temperature of the sample within shapes. This histogram shows the distribution of temperature from the drawn shapes across all of the events.

The next graph produced compares the number of layers contained within shapes to the frequency of those shapes. Most shapes had between 0 and 700 layers, but went up to over 900 layers per shape (Figure 8).

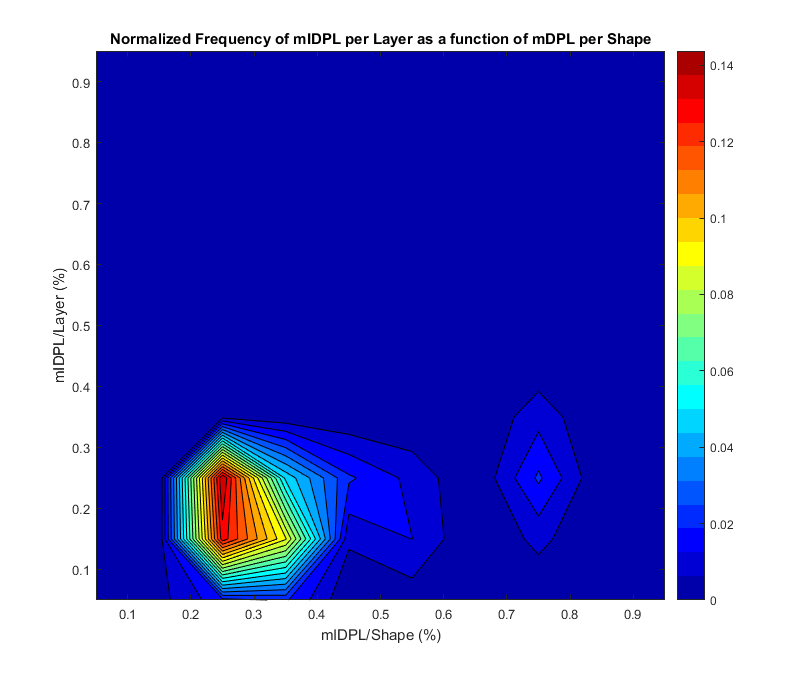
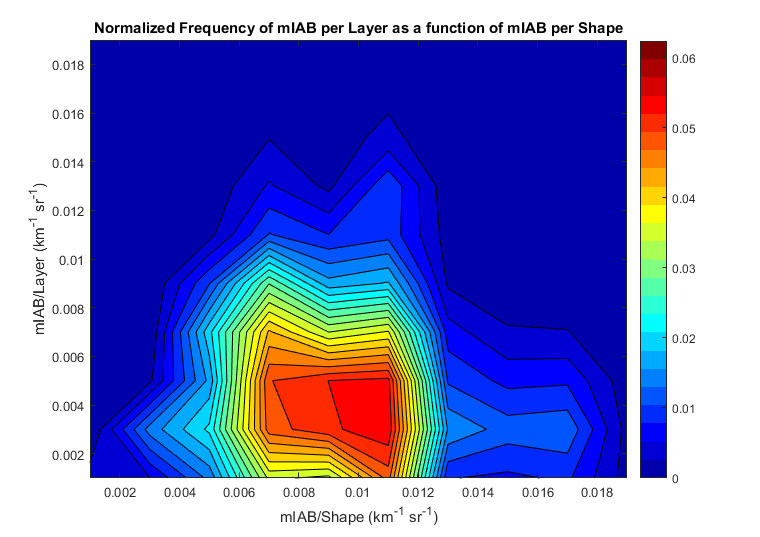


Finally, the last two graphs use contours to compare Level 1 data bounded by both draw shapes and Level 2 layers. The regions of the graph with a redder color indicate that layers or shapes are more frequent in that region. The x-axis of the graph indicates the averaged Level 1 value for each shape, and the y-axis indicates the averaged Level 1 value for each layer. Together, the plots show how the Level 1 data for drawn shapes and algorithmically determined layers differ.

*Figure 8:* Frequency of shapes vs the number of layers within shapes. This histogram shows the distribution of layers in drawn shapes across all of the events.

On the first contour plot (Figure 9), the total attenuated backscatter is compared. It shows that attenuated backscatter is concentrated around 0.006 sr-1 km‑1 to 0.012 sr-1 km‑1 for drawn shapes and 0.001 sr-1 km‑1 to 0.008 sr-1 km‑1 for layers. Therefore, the drawn shapes tended to have a slightly higher backscatter compared to layers. On the second contour plot (Figure 10), the depolarization ratio is compared. It shows that the depolarization ratio is concentrated at 25% for drawn shapes, while for layers it ranges from 15% to 25% percent. Therefore, the drawn shapes tend to have a much lower variance and slightly higher distribution compared to layers.

*Figure 9.* Normalized frequency of merged integrated backscatter per layer as a function of merged integrated backscatter per shape. The x-axis is the averaged backscatter attenuation for each shape and the y-axis is the averaged backscatter per layer as identified by the Level 2 data algorithms. The redder the contour, the more frequent shapes or layers are at that point.



*Figure 10:* Normalized frequency of merged integrated depolarization per layer as a function of merged integrated depolarization per shape. The x-axis is the averaged depolarization per shape and the y-axis is the averaged depolarization per layer. The redder the plot is at a given point, the more concentrated the shapes or layers are.

***4.3 Future Work***

Future work on this project would include further enhancements to make the software more intuitive, enhancements to the performance (especially panning speed), and eventually integration as a web application.

# 5. Conclusions

During the 2017 Summer term, the CALIPSO Cross-Cutting team integrated features from the previous term, added new features including tool integration for all levels of plots, persistent shapes, more dynamic import and export, a settings dialog and configuration file, and an improved drawing tool. The team created guidelines for version control, and updated documentation including instructions for both end users and future developers. The updates made this term were influence by end user recommendations and the teams own observations made when conducting the case study. These updates comprise the newest version of VOCAL, version 1.17.8.

These updates work together to make VOCAL easier for both the CALIPSO Science Team and external researchers to use. The added plots with tool integration allows users to view most of the data plots offered in the CALIPSO data products. Therefore, researcher will have greater access to the CALIPSO satellite data to enhance their research. The changes to the database backend will make it easier for researchers to share shapes drawn in VOCAL and enhance their ability to collaborate. Finally, the settings dialog and drawing tool make the VOCAL user experience smoother and more intuitive.

Because the tool is open source and hosted on a public development platform, researchers anywhere can access the tool and manipulate CALIPSO data. It therefor enhances NASA’s capacity building efforts in making Earth Observation data more useful to the public. CALIPSO data offers a unique view of the atmosphere that merits continued research. The enhancements to VOCAL created by the 2017 CALIPSO Cross-Cutting team improve on a tool that will aid in that research.

Finally, the case study produced a database of over 120 drawn shapes of user identified dust. Using our partners’ MatLab code we produced a number of plots that compare user-identified dust to that determined by the Level 2 data algorithms. Dr. Trepte was satisfied with the plots the team created, and plans to use similar analysis to help the CALIPSO science team improve CALIPSO data. Further, he will be able to use the updated VOCAL tool to continue to build databases of shapes to better understand CALIPSO data. Dr. Trepte expressed plans to use the tool in the fall to conduct more case studies to further their understanding of CALIPSO data.

# 6. Acknowledgments

Thank you to all of those who helped the CALIPSO Cross-Cutting Team this term:

* Dr. Charles Trepte (CALIPSO Science Team)
* Britney Hopgood (CALIPSO Science Team)
* Roman Kowch (CALIPSO Science Team)
* Dr. Kathleen Moore (NASA LaRC)
* Dr. Kenton Ross (NASA LaRC)
* Grant Mercer (UNLV)

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration.

This material is based upon work supported by NASA through contract NNL16AA05C and cooperative agreement NNX14AB60A.

# 7. Glossary

**Aerosol** – Small particles in the atmosphere from a number of sources; clouds, smoke, dust, and pollution are all aerosols

**Aerosol Subtype** – Part of CALIPSO’s Level Two data product, it indicates what type of aerosol a feature is including dust, smoke, marine, etc.

**Backscatter** – Light waves (electromagnetic radiation) that are reflected or scattered back towards the source when interacting with matter

**CALIPSO** – Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation

**Depolarization Ratio** – Part of CALIPSO’s Level One data product, it is the ratio of backscatter retuned in a perpendicular polarization orientation of the original laser beam to the total backscatter

**Git** – Version control software for software development. It allows users to track changes in code project’s history over time and integrate changes from multiple individuals

**GitHub** – An online remote repository service that offers integration with Git and allows users to publish open source code

**Horizontal Averaging** – Part of CALIPSO’s Level Two data product, it indicates in kilometers how large the horizontal sample is for a feature

**Ice/Water Phase** – Part of CALIPSO’s Level Two data product, it indicates if an area of the atmosphere is water, ice, or neither

**Layer ­**– Data from CALIPSO’s Level 2 product is organized as curtains of layers, or objects with averaged Level 1 data containing higher level info such as aeresol or cloud type

**MODIS** – MODerate resolution Imaging Spectroradiometer

**Steradian (sr)** – A three dimensional angle analogous to a radian, a conical section of a sphere. 1sr is a conical section with side length *r* and base area *r2.* It is used in the measure of attenuated backscatter

**Shape** – A user drawn polygon in VOCAL

**Total Attenuated Backscatter** – Part of CALIPSO’s Level One data product, a measure of how much light is backscattered from aerosols in units of sr-1 km‑1; higher total attenuated backscatter normally indicates denser aerosols

**Vertical Feature Mask** – Part of CALIPSO’s Level Two data product, it indicates general features in the atmosphere including (but not limited to) aerosol, cloud, and surface

# 8. References

Allen, B. (2015, July 31). Atmospheric Aerosols: What Are They, and Why Are They So Important? Retrieved June 23, 2017 from <https://www.nasa.gov/centers/langley/news/factsheets/Aerosols.html>

Falkowski, P., Barber, R., & Smetacek, V. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science, 281,* 200-206.

Liu, Z., Omar, A., Vaughan, M., Hair, J., Kittaka, C., Hu, Y. … Pierce, R. (2008). CALIPSO Lidar observations of the optical properties of Saharan dust: A case study of long-range transport, *Journal of Geophysical Research, 113*, D07207, doi:10.1029/2007JD008878.

Omar, A. Winker, D., Chieko, K., Vaughan, M., Liu, Z., Hu, Y., Hostetler, C. (2009). The CALIPSO Automated Aerosol Classification and Lidar Ratio Selection Algorithm*. Journal of Atmospheric and Oceanic Technology, 26*, 1995-2014.

Vaughan, M., Young, S., Winker, D., Powell, K., Omar, A., Liu, Z. … Hostetler, C. (2004). Fully automated analysis of space-based Lidar data: An overview of the CALIPSO retrieval algorithms and data products. *Proceedings of SPIE – The International Society for Optical Engineering*, 5575, November, 2004, doi:10.1117/12.572024.

Voiland, A. (2010, November 2). Aerosols: Tiny Particles, Big Impact. Retrieved August 7, 2017 from https://earthobservatory.nasa.gov/Features/Aerosols/

Winker, D., Pelon, J., & McCormick, M. (2006). Initial Results from CALIPSO. *Reviewed and Revised Papers Presented at the 23rd International Laser Radar Conference*, edited by C. Nagasawa and N. Sugimoto, pp. 991–994, Tokyo Metropolitan Univ., Tokyo.

Winker, D., Hunt, W., and McGill, M. (2007), Initial performance assessment of CALIOP, *Geophysical Research Letters* *34*, L19803, doi:10.1029/2007GL030135.

Winker, D. (2010, September). The CALIPSO Mission: A Global 3D View of Aerosols and Clouds [PDF]. Hampton: AMERICAN METEOROLOGICAL SOCIETY

Winker, D. (2016). CALIPSOLID L1 Standard HDF File - Version 4.10. NASA Atmospheric Science Data Center, accessed 26 June 2017. doi://10.5067/CALIOP/CALIPSO/LID\_L1-STANDARD-V4-10

Winker, D. (2016). CALIPSO LID L2 05kmMLay-Standard HDF File - Version 4.10. NASA Atmospheric Science Data Center, accessed 26 June 2017. doi://10.5067/CALIOP/CALIPSO/LID\_L2\_05KMMLAY-STANDARD-V4-10

Winker, D. (2016). CALIPSO LID L2 VFM-Standard HDF File - Version 4.10. NASA Atmospheric Science Data Center, accessed 26 June 2017. doi://10.5067/CALIOP/CALIPSO/LID\_L2\_VFM-STANDARD-V4-10