**NASA DEVELOP National Program**



NASA Langley Research Center

*Summer 2016*

CALIPSO Cross-Cutting

Enhancing the Usability of *Visualization of CALIPSO* (VOCAL)

Through a Test Case-Driven Approach

 **Technical Report**

August 11, 2016

Kathleen Moore (Project Co-Lead)

Ahmad Aburizaiza (Project Co-Lead)

Adama Ba

Andrea Martinez

Joseph Driscoll

Grant Mercer, University of Nevada, Las Vegas (Mentor)

Dr. Kenton Ross, NASA DEVELOP National Program (Science Advisor)

Previous Contributors:

Grant Mercer

Nathan Qian

Ashna Aggarwal

Jordan Vaa

Courtney Duquette

# 1. Abstract

The relationship that exists between cloud nuclei and aerosols plays a vital role in Earth’s climate system. For the purposes of detection, the Cloud-Aerosol Light Detection and Ranging (LiDAR) and Infrared Pathfinder Satellite Observations (CALIPSO) satellite offers scientists the ability to better understand the relationship between clouds, aerosols, and climate by measuring the backscatter created by embedded atmospheric aerosols. The standard visualizer for CALIPSO satellite data is limited in its extensibility by the proprietary programming language, Interactive Data Language (IDL), in which it is written. Hence, development of a new enhanced visualizer, Visualization of CALIPSO (VOCAL), has been in development by NASA DEVELOP teams since spring 2015. Written in Python, VOCAL displays CALIPSO curtain Hierarchical Data Format (HDF) files for both backscattered and depolarized Level 1-Version 3 data. In addition, the user can draw polygons overlaying the displayed data, to mark potential regions-of-interest. Subsequently, these regions can be tagged with attributes and notes. VOCAL can export to a database or JavaScript Object Notation (JSON) files any shapes that the user may want to save. During this particular term, the CALIPSO Cross-Cutting team added support for the newest CALIPSO data product (Version 4), alternative Graphical User Interface (GUI) options for the data in VOCAL, as well as a suite of use-cases for the software.

**Keywords**

Aerosol, VOCAL, CALIOP, LiDAR, HDFViewer, Open-Source

# 2. Introduction

* 1. ***Background Information***

Aerosols are small particles of liquid or solid suspended in the atmosphere. They influence Earth’s radiation budget by either limiting solar energy penetrating Earth’s atmosphere or by absorbing longwave energy released from the ground, causing the cooling or heating of the planet. There are several sources of aerosols, but those known to have the biggest effects on the climate are anthropogenic aerosols, volcanic plumes, and desert dust (Allen, 2015). All aerosols influence cloud properties; however due to the fact that anthropogenic aerosols have significant atmospheric concentration, they ultimately have the largest impact on the property of clouds to act as climate regulators (Forster et al., 2007; Wang et al., 2011). This influence has implications on surface temperatures and rainfall amounts globally (Ramachandran & Kedia, 2012). Aerosols can have a direct effect on human health as well. Unfortunately, combined effects of aerosols contribute to 7 million premature deaths worldwide (WHO, 2014). Understanding the origin and dispersion patterns of aerosols is imperative to understanding their true effect on climate and human health, and will allow decision-makers to improve policy implementation regarding emitted aerosol pollutants.

CALIPSO was launched in 2006 to record cross-sectional data of aerosols in the atmosphere, and it is the data from this satellite that is of particular interest to the CALIPSO Science Team at NASA Langley Research Center (LaRC). The purpose of the CALIPSO mission is to offer an alternative view of the interactions between aerosols and the atmosphere by offering a vertical display of atmospheric features up to 30 kilometers in altitude. The CALIPSO satellite probes the atmosphere with a Light Detection and Ranging (LiDAR) beam named “Cloud-Aerosol LiDAR with Orthogonal Polarization” (CALIOP) with wavelengths of 532 and 1064 nm, such that when the signal is returned to the satellite, the altitude and backscattering data of each particle within the atmosphere is recorded. The returned backscattering attenuation along with the depolarization ratio of the returned light distinguishes particle types and sizes, which can then lead to the classification of atmospheric particulate matter by subtyping regions as clouds, and the sources of which as natural or anthropogenic aerosols. The movement of the satellite and the quick pulses of the active LiDAR sensor create a large database, which requires visualization (Winker, 2007). The CALIPSO Science team has some automated methods for classification, but the team also values visual classification methods.

The end-user currently utilizes a CALIPSO data visualizer. After years of usage, the CALIPSO Science Team members had compiled various features that they would appreciate the visualizer having, but the software was written in IDL, an obscure proprietary programming language, making the editing of the code extremely difficult. Currently, after annotating regions of interest on the displayed satellite image, team members must then share this information in a disjointed fashion, separate from the annotations on the screen. To improve the way a CALIPSO data is displayed, DEVELOP teams have created the new VOCAL software to visualize the CALIPSO data from the original HDF into a GUI format.

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

**National Applications Addressed**

The CALIPSO Cross-Cutting project is unique in the sense that it inherently addresses several national application areas such as Health & Air Quality, Ecological Forecasting, Climate, and Weather. By using VOCAL to analyze CALIPSO data, scientists can establish point sources and dispersion patterns of pollutants to improve air quality. Along with health and air quality, VOCAL can improve the scientific understanding of clouds and aerosols, leading to a better understanding of how aerosols are impacting the climate and the weather. VOCAL does this by aiding in the understanding of cloud properties and relating their features to short timescales of precipitation variation along with larger impacts to a region's climatology.

**Project Partner**

The project partner is Dr. Charles “Chip” Trepte of the CALIPSO Science Team. The project partners will be using VOCAL to identify regions of interest within the atmosphere, share the denoted area, and perform statistical analysis on the region. VOCAL improves communication amongst the partners’ collaborators at other institutes and amongst each other. Most importantly, though, VOCAL can improve the user’s understanding of interactions between clouds and aerosols within the atmosphere.

**Project Objectives**

There are currently three project objectives for this project during term: evaluate the current version of VOCAL through rigorous use-case scenarios, enhance data visualization, and extend VOCAL’s compatibility to multiple satellite product levels and versions.

# 3. Methodology

This term’s instantiation of the project was meant to engage the feedback-loop process put forth at the conclusion of the previous term. Inherent in this phase of the project was continued software development as well as a user-centric evaluation of the experience of the software. The former objectives included the incorporation of previously existing goals for software functionality, as well as enhancements that would emerge from the user-evaluation during the current term. After completing a baseline set of use-cases, the members of the CALIPSO Cross-Cutting team addressed these issues concurrently, concluding with a fresh set of use-cases to establish goals for any future manifestations of the project.

VOCAL is software built using the Python programming language. It allows users to label the CALIPSO data with a polygon shape tool while modifying the displayed CALIPSO data through changes in latitude and altitude. When a shape has been drawn overlaying the HDF file on the visualization pane of the software, it can be tagged with attributes. The attributes are common aerosol subtypes and any supplementary notes that the user would like to add. The shapes then can be saved to a SQLite database or as a JSON file, allowing for future access to annotations and collaboration among scientists. Ultimately, the design of VOCAL tightly couples the shapes drawn by the user to the attributes that the user defines, pertaining to each of these shapes. These shapes can be saved, again, directly associated to the HDF file upon which they were originally drawn.

***3.1 Preliminary Testing***

One of the tasks that emerged from the conclusion of CALIPSO Cross-Cutting III was the need to rigorously test the software. While casual usage of VOCAL throughout software development has led to the discovery of different kinds of bugs, determining changes or enhancements to make to the software would only result from gathering accounts of the intended user experience of the software. Hence, the CALIPSO Cross-Cutting team began this term by acting as atmospheric scientists and using VOCAL like atmospheric scientists. Given different use-cases of tracking various kinds of aerosols all over the globe, each member of the team tracked a different target aerosol. These aerosols included: a West African dust plume, California wildfires, smoke from a Hawaiian volcano, and a Middle Eastern dust plume evolving into a semi-polluted dust plume as it traveled eastward. The team chose these scenarios based on the breadth of both the types of aerosols that they included and the geographic dispersion that they represented.

In addition to gaining nuanced background information on aerosols from having to research types of aerosols and their qualitative features within CALIPSO data, the team came away with several conclusions about the usability of the software. These conclusions were also supported by conversations with the project partner. First the visualizer did not include longitude information. The standard units of measure were altitude, latitude, and time, all depicted on the visualization pane, but longitude was not included. Second, while the software enables the user to export information about the polygons that were drawn and the underlying information to a database, and from the database to \*.txt or \*.csv file formats, no information about polygon coordinates was saved. Third, the tool used to draw polygons was found to be inaccurate at times resulting in multiple lines that did not enclose the selected region. Finally, the team determined that being able to visualize multiple kinds of data products from CALIPSO simultaneously would be useful in corroborating determinations on aerosol type.

***3.2 Software Development***

VOCAL is “living” software–in other words, it is in a functional state for basic annotation and exportation of CALIPSO satellite data, but it also exists in a state of evolution, to permit debugging and incorporation of new features. This term, the Cross-Cutting team addressed both existing bugs and enhancements that had been planned in the long-term, as well as those that arose from feedback acquired during the current term. Software development followed four distinct paths: enhancing VOCAL’s interface, data interpretation, data extraction, and compatibility.

***3.2.1 VOCAL Interface Enhancements***

In the previous version of VOCAL, the software interface was capable of viewing only one curtain data at a time. It was also limited to two curtain options, Level 1-Version 3 data: backscattered and depolarized. For example, a “Total Attenuated Backscattered” image is shown in Figure 1. For the end-user to analyze aerosols by comparing between two different data product curtains, the Cross-Cutting team determined that VOCAL should be able to render an alternate plot for visualizing the data.

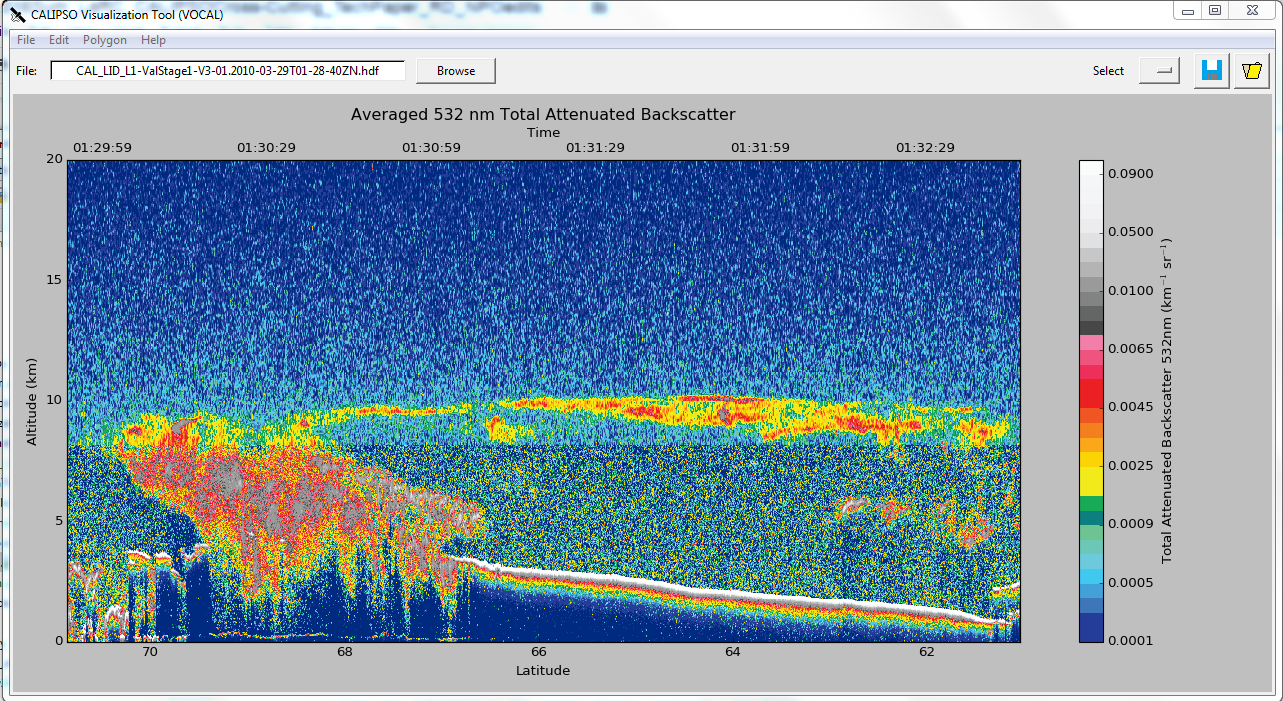


Figure 1. The “Total Attenuated Backscatter”Level 1 data product, as depicted within VOCAL.

The new version of VOCAL is implemented with a new Tkinter programmable object known as “ttk notebook.” Tkinter is the standard library for the Python programming language for use in building GUI components. The team employed methods from the ttk notebook to alter the interface to visualize the various CALIPSO curtains in a tabbable format as shown in Figure 2. Each tab represents a specific data curtain, for instance backscattered 532 nm, backscattered 1064 nm, virtual feature mask, and ice/water phase. It is more convenient for the end-user to see all CALIPSO curtains available in the HDF file for better interpretation and understanding of the aerosol features. In addition, a new menu is integrated into VOCAL to toggle the tabs. This menu helps the end-user to limit the number of shown tabs for more focused analysis.

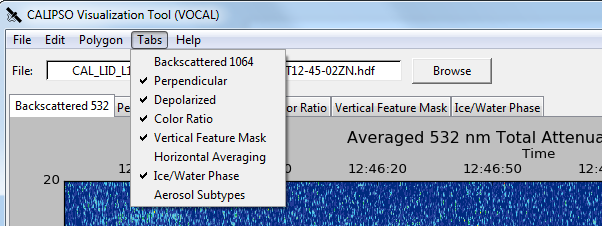


Figure 2. The new graphical user interface (GUI) in development for VOCAL. Instead of loading only one data product plot at a time, the user can now open a set of data products related to one HDF file using tabs, and the associated “Tabs” menu.

Another improvement added to VOCAL is the association of longitude along with latitude on the x-axis of the visualization area, as depicted in Figure 3.  Furthermore, VOCAL now exports the maximal extent of longitude, along with latitude, in the attributes of each polygon, delineating it in both of these spaces.

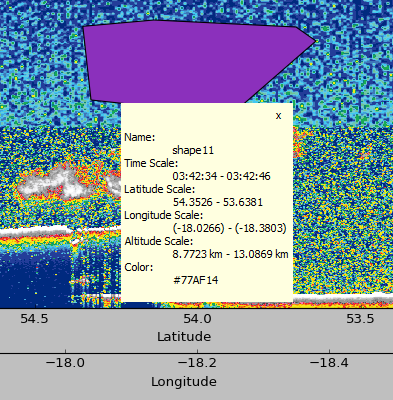


Figure 3. An example of the attribute pane for a polygon drawn in VOCAL. Note the addition of “Longitude Scale” in the pane as well as the secondary “Longitude” x-axis that have been added to the GUI.

Use-cases performed by both the Cross-Cutting team and by the end-user, Chip Trepte, highlighted the need for an important feature. Though VOCAL permits visualization of polygonal regions within its visualizer, it is often desirable to use other software for different spatial or statistical analyses. To then best incorporate information pertaining to the polygons, it would be necessary to export the coordinates of each vertex of each polygon. VOCAL now enables the export of this list of coordinates from the database into \*.txt and \*.csv files, in terms of (*time*, *altitude*) coordinate pairs, as shown in Figure 4.  Dr. Trepte confirmed that this format would be most useful for his current needs.

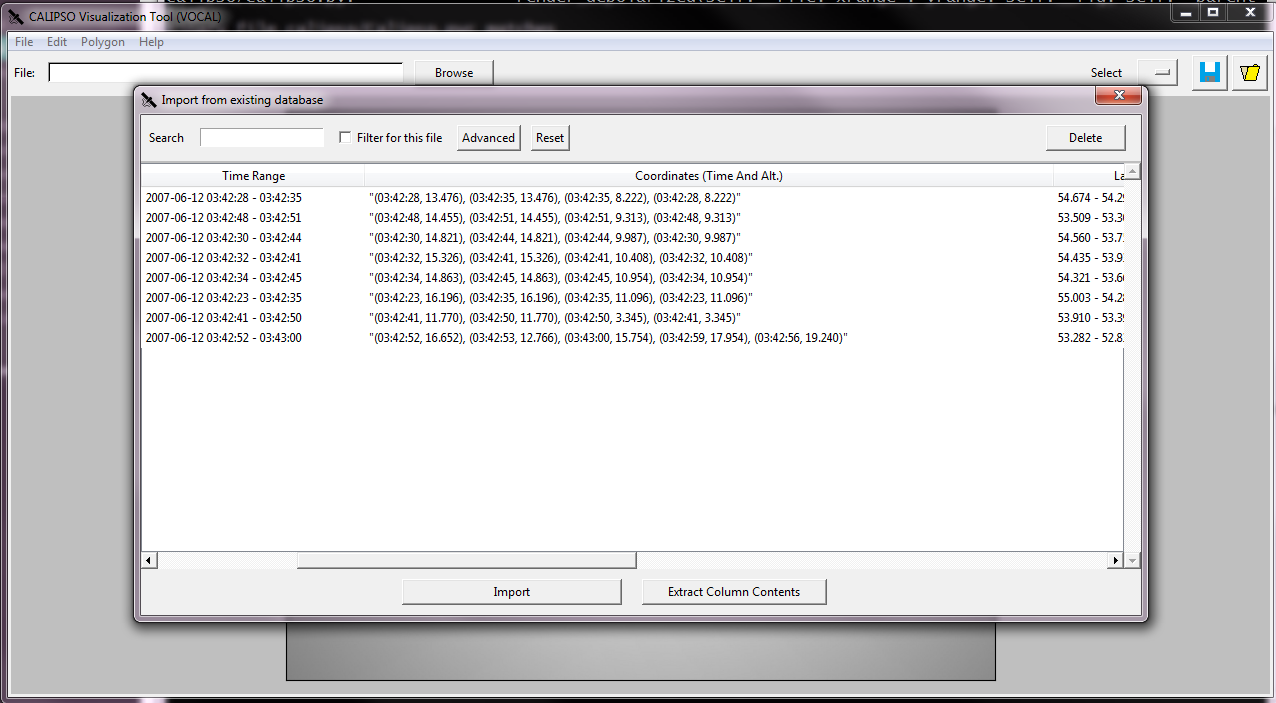


Figure 4. The interface for importing from the local database in VOCAL is shown with the added polygon coordinate lists added.  These coordinates can also be further exported to \*.txt and \*.csv files for future analyses.

Annotating the aerosol features in the previous version of VOCAL was difficult because an end-user was not able to see a floating line while drawing to track the edges of an aerosol feature. To close a polygon, the last segment **must** cross the first segment, which reduced the accuracy of an aerosol feature boundary. In this term, a new mechanism of annotating the aerosols was developed. A floating line helps the end-user to draw along a specific aerosol features more accurately. Also, when the end-user wants to draw the last segment to close the polygon, the user can simply click the right-mouse-button and the last segment will snap to the first point of the first segment automatically.

***3.2.2 Data Interpretation***

At the start of the term, VOCAL displayed Level 1 CALIPSO satellite data in two formats: total attenuated backscatter and the depolarization ratio, both for the 532 nm wavelength LiDAR pulse. Total attenuated backscatter represents the proportion of the signal that returns as it is reflected off of aerosols in the atmosphere.  Often depicted in false-color, a typical backscatter image describes a range of attenuation based on the different particles encountered, from zero backscatter to full backscatter. Depolarization ratio helps to identify the shapes of the particles. Depending on the value, the ratio indicates a more spherical or a more irregularly-shaped particle, which in turn has an effect on the direction of the polarized light that is returned to the CALIOP sensor.

Because each output CALIPSO HDF file is a three-dimensional array comprised of numerous layers (of which total attenuated backscatter and depolarization ratio are a part), each layer provides a different analysis of the underlying backscattering data. The team provided extendibility to alternate layers—those that are Level 1, more rasterized, per-pixel backscattering data, to Level 2, more abstracted, polygonal regions of classification, including the “Vertical Feature Mask (VFM)” (Figure 5), and “Ice/Water Phase (IWP)” (Figure 6) data products. Having access to more visualization styles can further refine scientists’ intuition about potential aerosol plumes.

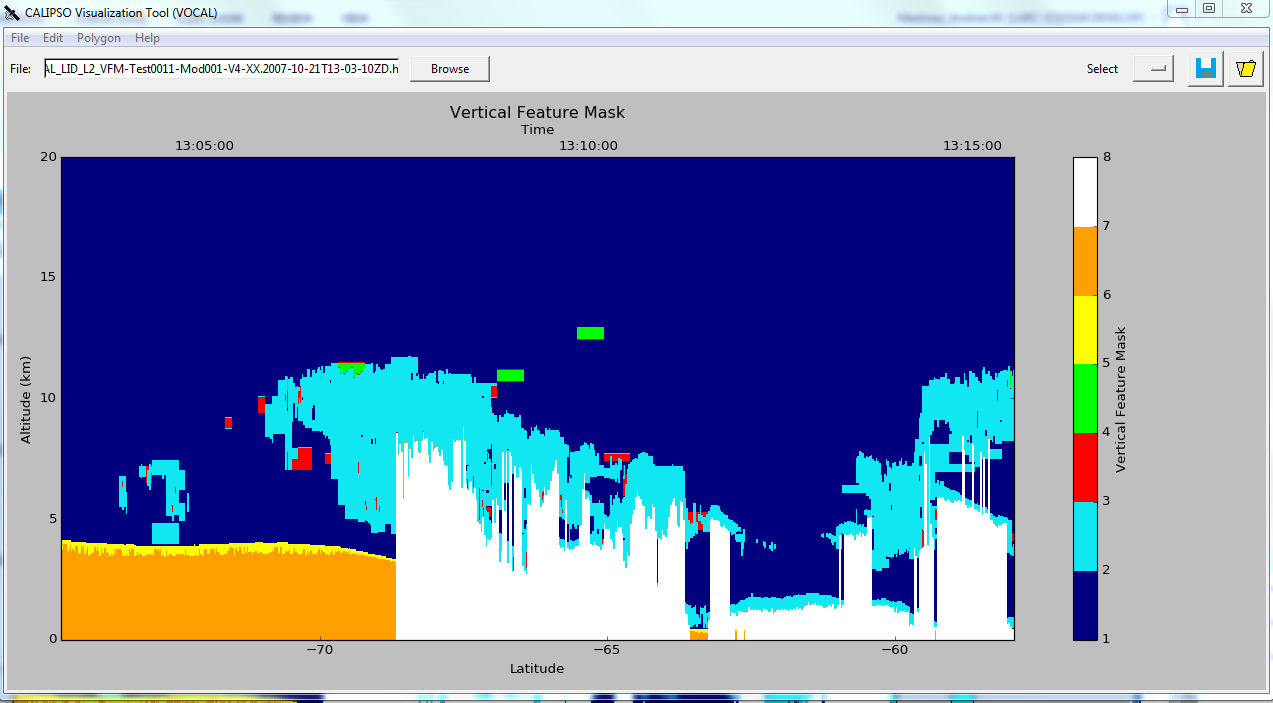


Figure 5. The “Vertical Feature Mask” Level 2 data product from CALIPSO as depicted within VOCAL.

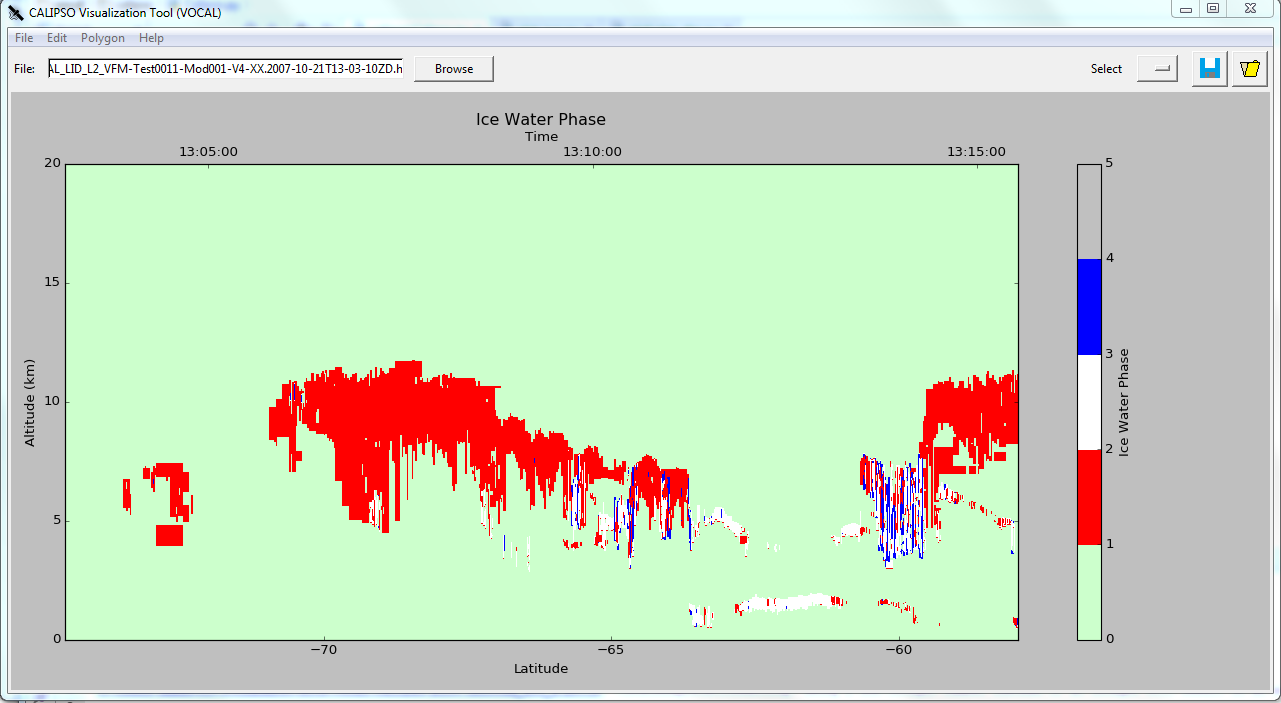


Figure 6. The “Ice Water Phase” Level 2 data product from CALIPSO as depicted within VOCAL.

***3.2.3 Back-end Refactoring***

Until the current term, the code underlying the importing and manipulation of CALIPSO data treated each data product separately, often duplicating code unnecessarily and disassociating data products even if they are related within the same level. This term, back-end code has been refactored, treating CALIPSO data as a “data block.” For a comparison of these pre-existing and new storage options, refer to Figure 7. On NASA servers, CALIPSO HDF data is treated as a three-dimensional block, at least within a data product level, such as Level 1. This idea is now mirrored within the code of VOCAL, essentially treating the HDF data as a three-dimensional array from which “slices” are extracted, one per data product, for display. Once a data product is requested by the user for visualization, it persists. Then, because of the addition of the tabbed interface as discussed in section 3.2.1, subsequent data products can also be loaded in separate tabs. Previous data remain unless a tab is closed.

Figure 7. Left: Data organization pre-term. From the HDF file, backscattered and depolarized Level 1 data products are loaded separately in two different methods. Right: Data organization post-term. Once one data product, i.e. Backscattered, is loaded, it is stored in the “VOCALDataBlock” data structure. Subsequent data products are also loaded into this data structure, and there is support for essentially numerous data products from both Levels 1 and 2.

***3.2.2 Compatibility***

In CALIPSO Version 3 data, Level 1 curtains are stored in a separate file from Level 2 curtains. The reader-files for the version of VOCAL from the previous term were implemented with regards to reading only Level 1 of Version 3 data. This term, the Cross-Cutting team successfully visualized Level 2 curtains of Version 3 data with their implementation of the VFM and IWP data products. In addition, the team gained early access to Version 4 data (not yet publicly available) to test for compatibility with the software. The team subsequently ensured that VOCAL could import Level 1 and Level 2 curtains from Version 4 data in addition to Version 3 data.

# 4. Results & Discussion

***4.1 Analysis of Results***

Each team member ran a specific use-case scenario for preliminary evaluation of the software. The use-cases showed flaws and strengths in the VOCAL software. The test cases emphasized the limitation of the software both in visualizing multiple files at once and performing qualitative testing on the defined shapes through statistical analysis. VOCAL was also tested to ensure all versions of CALIPSO data could be opened using the software.

Due to the subsequent increased usability and functionality of the program from the aforementioned enhancements, the Cross-Cutting team ran another set of use-case studies to reevaluate the state of the software. For instance, one team member compared selections made on Level 1 attenuated backscatter data to those automatically generated in the Level 2 vertical feature mask. The Cross-Cutting team member was immediately able to compare the estimation to those made algorithmically, by the CALIPSO Science Team.

As these new features are merged into the master copy of VOCAL for distribution, the features will continue to be evaluated for usability and effectiveness. Because the end-user, Dr. Trepte, will have access to these new features, the ideal, intended user will indeed be in the process of evaluation. One limitation of this burst of new features is ensuring that no new bugs are introduced and that the features are inter-compatible. Thus, the code-merging process will continue through the official end of the term.

***4.2 Future Work***

VOCAL can continue to be improved to better fit the end-user’s needs. There are multiple features that can be added to the current software. A feature that has been discussed is the addition of a ‘magic wand’ tool in lieu of either of the polygon shape drawing tools discussed. In this case, the user would only need to touch the center of what the user believes is a target aerosol region. The system would then automatically ‘inflate’ from this point to the most logical borders of the region, semi-automatically selecting the target region. Another feature that can be added is the ability to export the annotated polygons of the aerosol features into formats readable by digital globes such as NASA World Wind, Google Earth, and ArcGIS Earth.

We also anticipate that VOCAL will evolve from a desktop to a web based application with digital globe visualization capabilities implemented in Google Earth Engine, Cesium, or the D3.js library. Having VOCAL as a web application will make it cross-platform since it will run through an internet browser. In addition, the annotated aerosol features can be stored in a shared online database such as PostGIS/PostgreSQL. There also should be a control on the data access based on a user’s granted privileges.

# 5. Conclusions

In this term of CALIPSO Cross-Cutting, the project team re-evaluated VOCAL through use-case scenarios, subsequently using the results of these evaluations to inform decisions on software development. Key to this motivation is the idea that the team would use the software like the intended users: atmospheric scientists. New features added to the software this term include: a tabbed visualizer allowing the user to see multiple data products at once, more verbose display information for the data plot and for polygons, and back-end data storage reconfiguration to facilitate the simultaneous display of multiple data products. Additionally, the team ensured, specifically, that Level 2 and Version 4 data products would be supported in VOCAL, as the CALIPSO Science Team is moving towards better confidence in these versions of the data products. This feature also allows scientists more flexible means for the inspection of atmospheric properties in CALIPSO curtain data. To further ensure flexibility, the team added more verbose polygon annotation identification and export capabilities by integrating longitude range into polygon attributes and to the display, and by adding vertex coordinate lists to the attributes of each polygon to facilitate viewing polygons in other geospatial programs.

The end-user, Dr. Charles Trepte, is satisfied that VOCAL and the incorporation of these new features make the software a legitimate means of validation for the algorithms used by the CALIPSO Science Team to identify aerosols in CALIPSO curtain data. For the user in general, the combination of these modifications will improve the end-user experience by enhancing the usability of VOCAL and increasing the breadth of support for various CALIPSO data products.

# 6. Acknowledgments

* Grant Mercer, NASA DEVELOP LaRC, Summer 2015, Team Lead
* Dr. Kenton Ross, NASA DEVELOP National Program Science Advisor
* Michael Bender, Technical Lead, NASA DEVELOP National Program
* Jordan Vaa, IT Fellow, NASA DEVELOP National Program
* Emily Gotschalk, Center Lead, NASA DEVELOP National Program
* Tyler Rhodes, Center Lead, NASA DEVELOP National Program
* Carrie Kelley, Assistant Center Lead, NASA DEVELOP National Program

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 NNL11AA00B and cooperative agreement NNX14AB60A.

# 7. References

Allen, B. (2015, July 31). Atmospheric Aerosols: What Are They, and Why Are They So Important? Retrieved June 23, 2016 from https://www.nasa.gov/centers/langley/news/factsheets/Aerosols.html

Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D., Haywood, J., Lean, J., Lowe, D., Myhre, G., Nganga, J., Prim, R., Raga, G., Schulz, M., & Dorlan, R. (2007). Changes in Atmospheric Constituents and in Radiative Forcing in Climate Change: The Physical Science Basis. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Liu, Z., Omar, A., Vaughan, M., Hair, J., Kittaka, C., Hu, Y., Powell, K., Trepte, C., Winker, D., Hostetler, C., Ferrare, R., 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(D7).

NASA. (2014, December 31). CALIPSO Data User's Guide - Essential Reading. Retrieved from: <http://www-calipso.larc.nasa.gov/resources/calipso_users_guide/essential_reading/index.php>

Ramachandran, S. & Kedia, S. (2012). Aerosol, clouds, and rainfall: inter-annual and regional variations over India. *Climate Dynamics,* 40(7), 1591-1610.

Swap, R., Garstang, M., Greco, S., Talbot, R., & Kållberg, P. (1992). Saharan dust in the Amazon Basin. *Tellus B*, 44(2), 133-149.

World Health Organization (WHO). (2014, March 25). 7 million premature deaths annually linked to air pollution. Retrieved from: <http://www.who.int/mediacentre/news/releases/2014/air-pollution/en/>

Wang, M., Ghan, S., Ovchinnikov, M., Liu, X., Easter, R., Kassianov, E., Qian, Y., & Morrison, H. (2011). Aerosol indirect effects in a multi-scale aerosol-climate model PNNL-MMF. *Journal of Atmospheric Chemistry and Physics.* 11, 5431-5455.

Winker, D. (2007). The CALIPSO Mission and Initial Observations of Aerosols and Clouds from CALIOP. *Optics InfoBase Conference Papers.*

# 8. Content Innovation

**Content Innovation #1**

Glossary Viewer:

* CALIOP: Cloud-Aerosol LiDAR with Orthogonal Polarization, one of three sensors aboard the CALIPSO satellite
* CALIPSO: Cloud-Aerosol LiDAR and Infrared Pathfinder Satellite Observation, a satellite launched in 2006 as a joint venture between NASA and CNES, the French government space agency
* Depolarization ratio: the quotient of cross-polarized to parallel-polarized deviation angles from the polarization orientation of the original input signal
* Total attenuated backscatter: the coefficient that represents the attenuation of the input signal from the backscattered direction
* VOCAL: Visualization of CALIPSO, a visualizer for the CALIPSO satellite data products written in the Python programming language

**Content Innovation #2**

Inline Supplementary Material

* Figure 1. The “Total Attenuated Backscatter” Level 1 data product, as depicted within VOCAL.
* Figure 2: The new graphical user interface (GUI) in development for VOCAL. Instead of loading only one data product plot at a time, the user can now open a set of data products related to one HDF file using tabs, and the associated “Tabs” menu.
* Figure 3. An example of the attribute pane for a polygon drawn in VOCAL. Note the addition of “Longitude Scale” in the pane as well as the secondary “Longitude” x-axis that have been added to the GUI.
* Figure 4. The interface for importing from the local database in VOCAL is shown with the added polygon coordinate lists added.  These coordinates can also be further exported to \*.txt and \*.csv files for future analyses.
* Figure 5. The “Vertical Feature Mask” Level 2 data product from CALIPSO as depicted within VOCAL.
* Figure 6. The “Ice Water Phase” Level 2 data product from CALIPSO as depicted within VOCAL.
* Figure 7. Left Data organization pre-term. From the HDF file, backscattered and depolarized Level 1 data products are loaded separately in two different methods. Right: Data organization post-term. Once one data product, i.e. Backscattered, is loaded, it is stored in the “VOCALDataBlock” data structure. Subsequent data products are also loaded into this data structure, and there is support for essentially numerous data products from both Levels 1 and 2.

**Content Innovation #3**

Video Tutorial (Audio Slides) of Software Usage

Shared through Google Drive at: <https://drive.google.com/file/d/0BwJP7qDvPfz1ZEtwN2ZfN1BYdDQ/view?usp=sharing>

**Content Innovation #4**

VPS in English and another version in French

Shared through Google Drive at:   
English version: <https://drive.google.com/open?id=0BxaeN9dqbfQ4RkFFR3Fqcmk2SEk>  
French version: <https://drive.google.com/open?id=0BwJP7qDvPfz1enZnMGVMVGtrajQ>

# 9. Appendices

VOCAL Documentation website: <http://syntaf.github.io/vocal/>