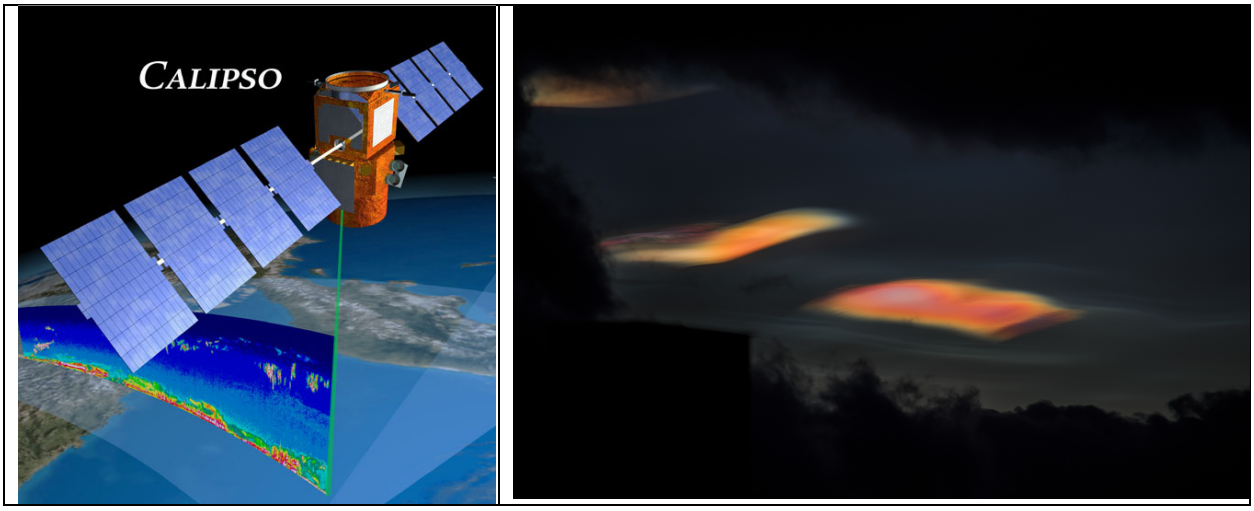
Comparison of Polar Stratospheric Cloud (PSC) Data from the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) sensor on the joint NASA-CNES CALIPSO Satellite

By Walt Baskin, SSAI Senior Software Engineer, NASA Atmospheric Science Data Center Aug, 2019

This Applied Data Science Capstone project will use data science techniques covered in this course to prepare and analyze PSC observations using LIDAR sensor data from two independent data products produced from the CALIPSO satellite mission archived at the NASA Langley Research Center's Atmospheric Science Data Center (ASDC).



Left: Depiction of the CALIPSO Satellite with a vertical profile 'LIDAR data curtain' (credit: NASA)
Right: Polar Stratospheric Clouds over Oslo Norway (credit: Gabriel Arne Hofstra)

1. A description of the problem and a discussion of the background.

PSCs form at higher latitudes (greater than 50deg latitude in the Arctic and less than -50deg latitude in the Antarctic) in the polar winter at altitudes above the tropopause and contribute to ozone depletion in the stratosphere. The Wikipedia article on PSCs (https://en.wikipedia.org/wiki/Polar_stratospheric_cloud) is a good source of information describing this phenomenon. The CALIOP LIDAR sensor on the CALIPSO satellite particularly well suited for observing and classifying PSCs. Details of the CALIPSO mission can be found at the following URL: https://www-calipso.larc.nasa.gov/.

The CAL_LID_L2_PSCMask product focuses on Polar Stratospheric Cloud observations and provides detailed classifications of different types of PSCs and measurements that are useful to PSC researchers for analysis of the physical and chemical composition of PSCs. The CAL_LID_L2_VFM product provides more generalized classifications of all types of

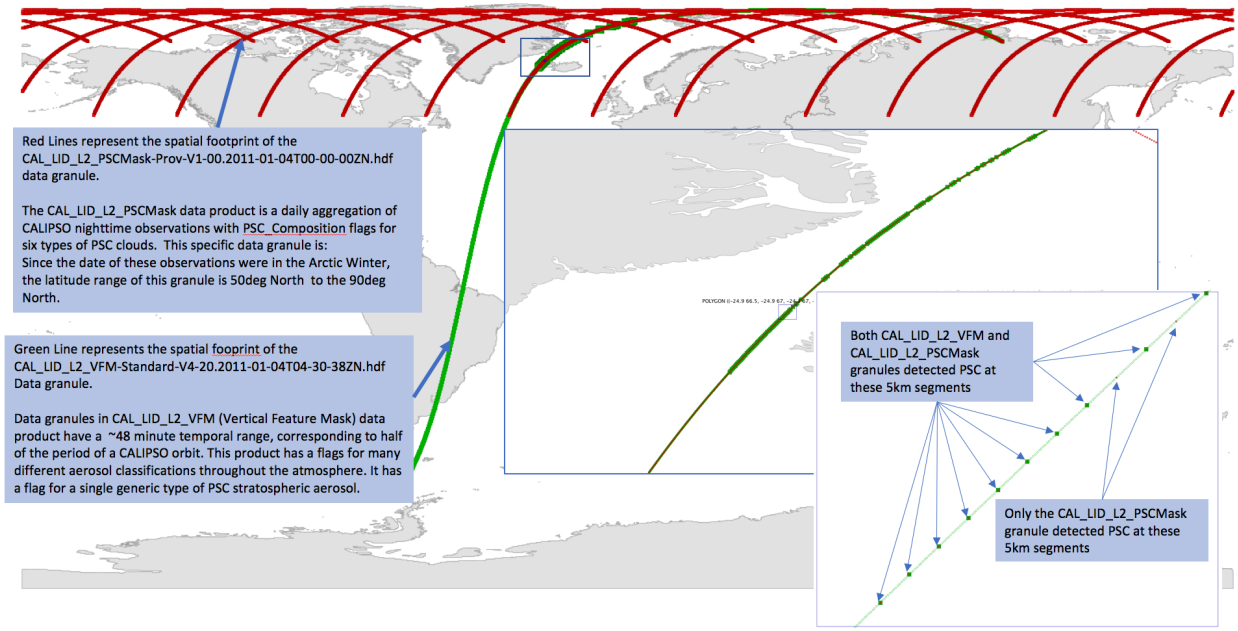
atmospheric aerosols and is used by a much larger community of atmospheric scientists. The algorithm used to generate the PSC Aerosol Sub-type flag in the VFM product was developed independently from the algorithms used in the PSCMask Product. A correlation analysis between the PSC data in these two products will help members of the ASDC and CALIPSO science team to determine the specific types of PSC phenomenon best represented by the CAL_LID_L2_VFM product and will help guide decisions on modifying the PSC detection algorithms in both products.

This project will harvest one month of data from two satellite data products (CAL_LID_L2_PSCMask and CAL_LID_L2_VFM) depicting PSC observations averaged at 5km intervals along the CALIPSO Satellite's LIDAR sensor path and will perform regression analyses comparing classifications of various types of PSCs. The data will be accessed and subsetted through the ASDC OPeNDAP server leveraging the Python pydap and pyhdf libraries. URLS for the specific month of data to be processed are listed below:

ASDC OPeNDAP URLS to access CALIPSO data granules for his analysis (Jan 2011)
https://opendap.larc.nasa.gov/opendap/CALIPSO/LID_L2_PSCMask-Prov-V1-00/2011/01/contents.html
https://opendap.larc.nasa.gov/opendap/CALIPSO/LID_L2_VFM-Standard-V4-20/2011/01/contents.html

2. A description of the data and how it will be used to solve the problem.

The image below illustrates the spatial coverage of the two data products that will be used in this analysis.



Preparation of CAL_LID_L2_PSCMask Data

Data arrays from seven HDF Science DataSets (L1_Input_Filenames, Latitude, Longitude, Profile_Time, PSC_Composition, Tropopause_Height, Orbit_Index) will be read from each daily CAL_LID_L2_PSCMask granule and split into ~15 Pandas dataframes (each dataframe corresponds to a single orbital pass). The PSC_Composition array consists of integer values ranging from 0 to 6. Seven columns (compflag1 to compflag7) will be added to each CAL_LID_L2_PSCMask Pandas DataFrame. Values in each of these compflag# columns are set to 1 if any flag is detected in the vertical column corresponding to the specific spatial-temporal observation. See the table below for a breakdown of types of PSC_Composition flags mapped to the compflag# column values.

PSC_Composition Value	Pandas Column Name	Type of PSC Observed
0		No Cloud detected
1	compflag1	Liquid Supercooled Ternary (sulfuric acid, water, nitric acid) Solution (STS) droplets
2	compflag2	Mix1: STS + low number densities/volumes of Nitric Acid Trihydrate (NAT) particles
3	compflag3	Mix2: STS + intermediate number densities/volumes of NAT particles
4	compflag4	Water ice clouds
5	compflag5	Mix2-enhanced: STS + high number densities/volumes of NAT particles
6	compflag6	Wave ice: Mountain wave induced water ice clouds (R > 50)
	compflag7	Any PSC_Composition value > 0

Preparation of CAL_LID_L2_VFM Data

Spatial subsets (Latitudes >= 50Deg North) of data arrays from four HDF Science DataSets (Latitude, Longitude, Profile_Time, Feature_Classification_Flags) will be read from the night half-orbits of each CAL_LID_L2_VFM granule, into a Pandas dataframe. Two bit masks must be applied to the two byte Feature_Classification_Flags value to determine if a specific Feature Type (bits 1-3) and Sub-Type (bits 10-12). This project will focus on Feature Type: stratospheric aerosol (4) and Feature Sub-Type: PSC aerosol (1). The following table depicts the some of the bit masks for determining Feature Type and Sub-Type (see the CALIPSO Data Products Catalog (DPC) for a full description of the Feature_Classification_Flags bit masks. URL: https://www-calipso.larc.nasa.gov/products/CALIPSO_DPC_Rev4x20.pdf)

Bit(s)	Field Description	Bit Interpretation
1-3	**Feature Type**	**0 = invalid (bad or missing data)**
		1 = “clear air”
		2 = cloud
		3 = tropospheric aerosol
		4 = stratospheric aerosol
		5 = surface
		6 = subsurface
		7 = no signal (totally attenuated)
10-12	**Feature Sub-type**	

If feature type = cloud, bits 10-12 will specify the cloud type.	0 = low overcast, transparent	
		1 = low overcast, opaque
		2 = transition stratocumulus
		3 = low, broken cumulus
		4 = altocumulus (transparent)
		5 = altostratus (opaque)
		6 = cirrus (transparent)
		7 = deep convective (opaque)
If feature type = Stratospheric Aerosol, bits 10-12 will specify the stratospheric aerosol type.	0 = invalid	
		1 = PSC aerosol
		2 = volcanic ash
		3 = sulfate/other
		4 = elevated smoke

The value of the Pandas DataFrame column labeled 'vfm_flag' will be set to 1 if any vertical profile in the Feature_Classification_Flags array corresponding to the 5km spatial-temporal observation contains a PSC aerosol Feature Sub-type.

Correlation Analysis

A Pearson's correlation method will be used to calculate correlation coefficients for each orbit comparing the vfm_flag values from CAL_LID_L2_VFM data with the seven compflag# values from the CAL_LID_L2_PSCMask data. Correlation R Values will be summarized by appropriate plots using the Python matplotlib library.