Simplifying Satellite and Ground Data Validation with Level-2 Subsetting

Poster: A33L-2957

NASA/Goddard EARTH SCIENCES DATA and INFORMATION SERVICES CENTER (GES DISC)

Johnson, James[1,2], Paul Huwe[1,2], David S. Silberstein[1,2], Jerome Alfred[1,2], Jennifer Adams[1,2], Dana Ostrenga[1,2]

[1]NASA Goddard Space Flight Center, [2]ADNET Systems Inc.,
Contact: paul.huwe@nasa.gov

Abstract

We demonstrate how scientists can simplify their satellite data validation workflow with the use of NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) subsetting services. We perform a sample validation of ozone products co-located with ground-based total ozone measurements using subsetting services to trim satellite data to only the relevant user-defined variables and spatiotemporal region. The subsetting service automatically returns only relevant data granules that adhere to a set of user-defined coincidence criteria, which greatly reduces user workload. Moreover, the resultant data files are substantially smaller than full data granules due to the subsetting service further culling the data to the relevant spatiotemporal coincidence criteria, user-defined variables, and user-defined dimensions of variables. This decreases data download throughput and file storage requirements. The validation presented here quantifies the time and file size savings that can be achieved by utilizing subsetting services within the satellite data validation workflow.

Data Selection

Ground Validation Sites

We selected three stations that measure total ozone: Utqiaġvik (Barrow) Alaska, Haute Provence in France, and the South Pole Observatory. These stations use Dobson Ozone Spectrophotometers to measure the total column of atmospheric ozone. Data are available from the NOAA Earth System Research Laboratory Global Monitoring Division (ESRL/GMD).

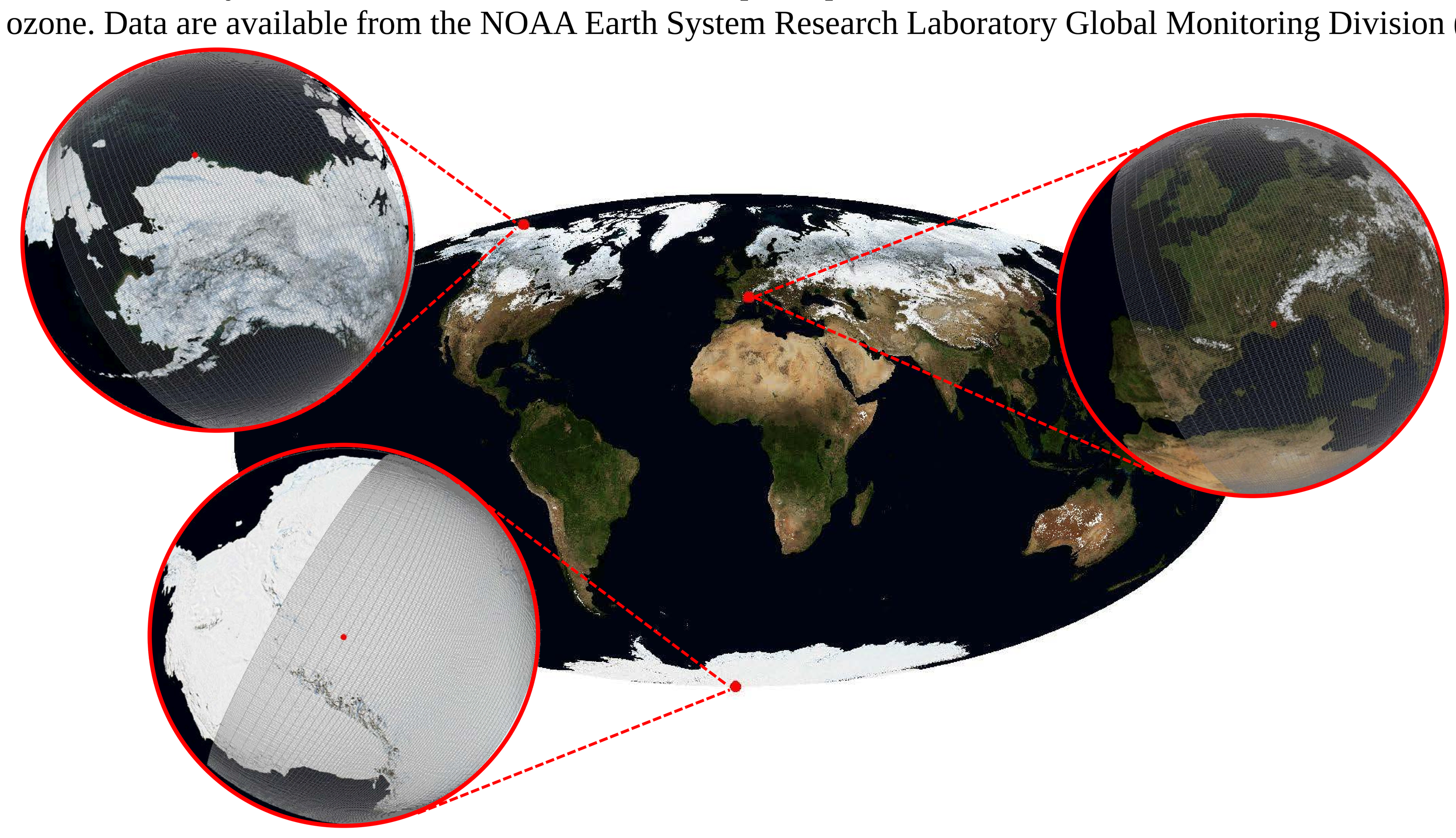

Figure 1. A global map of the Earth showing the locations of our selected ground-based observation stations. Insets zoom in on the area with an overlay of an OMI swath in grey and a red dot at the station location.

Satellite Measurements

We chose four satellite-borne retrievals of total column ozone: two Ozone Monitoring Instrument (OMI) products and one Microwave Limb Scanner (MLS) product from the Aura satellite, as well as a Tropospheric Monitoring Instrument (TROPOMI) product from the Copernicus Sentinel 5 Precursor (S5P) satellite.

	OMI-TOMS	OMI-DOAS	TROPOMI	MLS
Spatial Resolution	13 km x 24 km	13 km x 24 km	7 km x 3.5 km	165 km x 3 km
Data Granule Period	98 minutes	98 minutes	101.5 minutes	1 day
Daily Coverage	Global	Global	Global	82°S to 82°N
Data Granule Size	~35 MB	~15 MB	~250 MB	~1.4 MB

Table 1. A summary of the satellite measurement characteristics.

Methodolog y

We use the GES DISC Application Program Interface (API) to automate the satellite data co-location and download. Python scripts use the API interface to search through the Common Metadata Repository (CMR) for coincident data granules and then invoke the Level-2 subsetting service to extract the closest data point within a 300 km radius from the ground-based observation stations. The service is further tailored to use variable selection in order to cull the data to only the variables of interest. The use of the CMR search and the subsetting service streamline the workflow and minimize user interaction.

Workflow Components

- Select a data parameter (e.g. Total Ozone)
- Create a list of the satellite platforms and products that are measuring the selected data parameter
- Choose ground-based observations stations that have a long record of selected data parameter measurements
- API scripts find the satellite data granules that overlap the stations and are co-located in time with the observations
- Decide which are the relevant variables to save from each data granule
- API scripts invoke the Level-2 subsetting service to extract the variable and spatial subset from each granule and then download the output data files
- Consolidate all the data into a preferred analysis and visualization tool (e.g. IDL, GrADS, Python, etc.)
- Perform the comparisons and report on the results

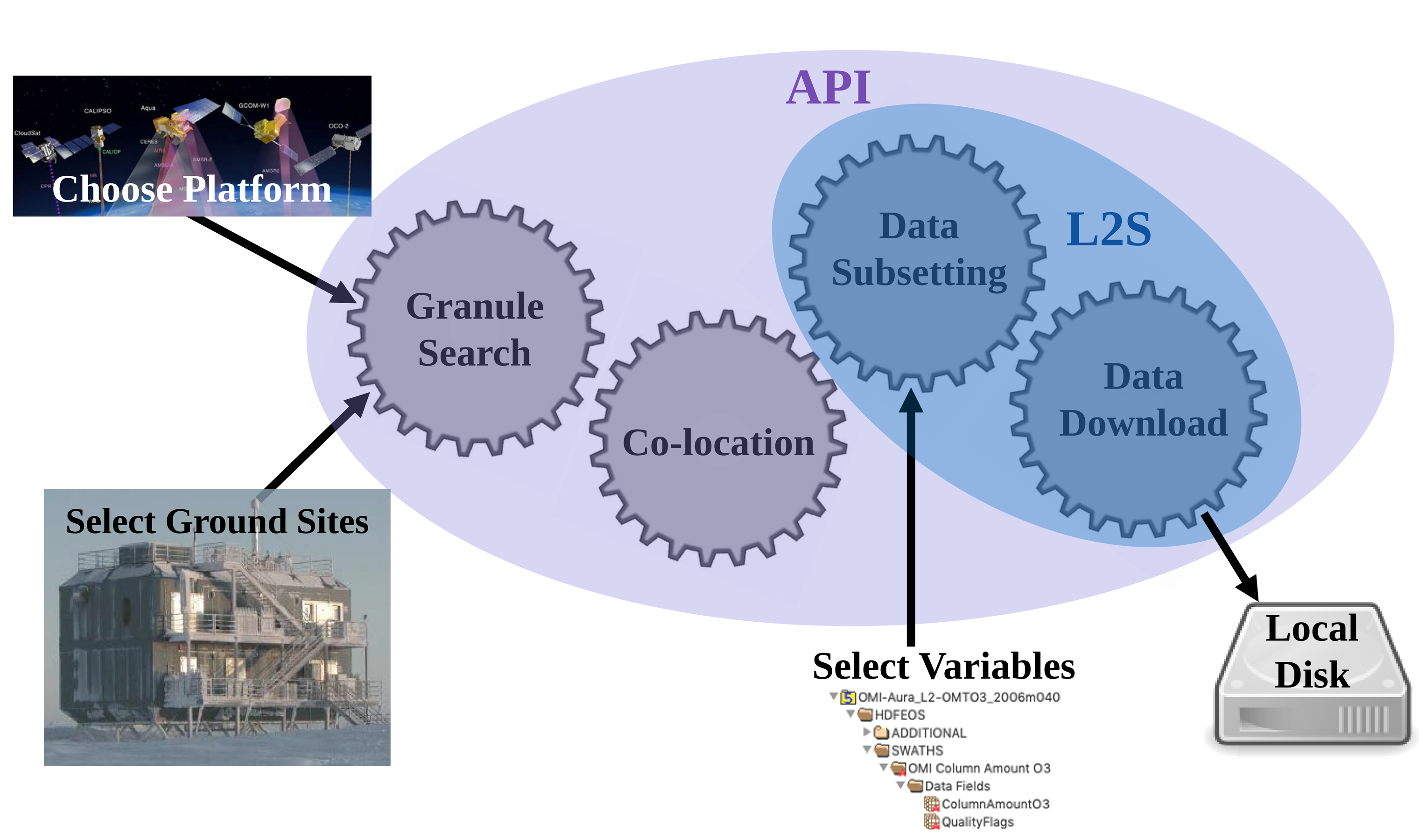

Figure 2. An illustration of the workflow that uses API scripts to invoke the Level-2 Subsetting (L2S) service and download data subsets.

Results

The scatter diagrams in Figure 3 illustrate the agreement between the ground-based and satellite-borne ozone measurements. Each dot represents a co-located pair of data points, with the perfect correlation line overlaid in black. The MLS product shows a low bias because the limb scanner does not measure to the surface (below 261 hPa) and misses most of the tropospheric ozone. The OMI measurements at Utqiaġvik and Haute Provence show more scatter and bias due to cloud contamination; correlation is better at the South Pole where there are fewer clouds. Correlation statistics are in Table 2.

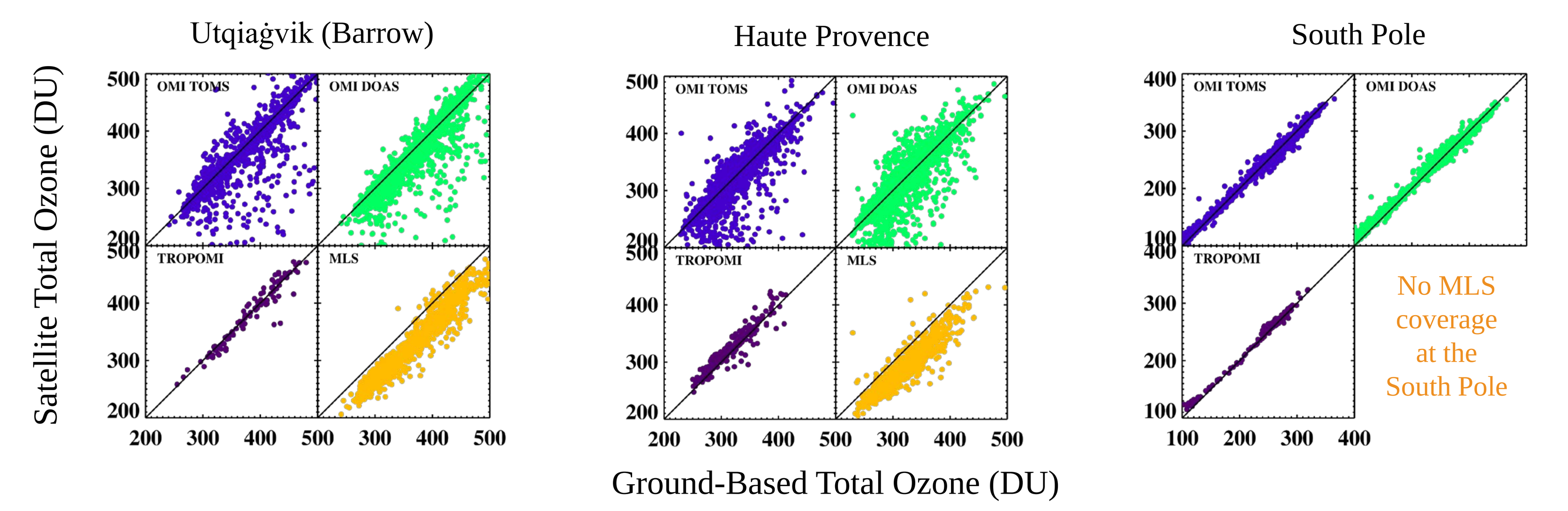

Figure 3. Scatter plots of satellite total ozone vs. ground-based ozone measurements.

	OMI-TOMS		OMI-DOAS		TROPOMI		MLS	
Station	Slope	R^2	Slope	R^2	Slope	R^2	Slope	R^2
Utqiagvik (Barrow)	0.979	0.232	0.981	0.791	1.052	0.968	0.977	0.922
Haute Provence	0.880	0.244	0.949	0.635	0.941	0.905	0.879	0.838
South Pole	0.984	0.993	0.961	0.992	1.002	0.992		

Table 2. A summary of the linear regression fits of the satellite measurements to the ground-based measurements. The parameters shown are the Slope and Correlation Coefficient (R^2).

The time series plots in Figure 4 are from the Haute Provence location and span the period from October 2004 to present. The top panel shows the observed total ozone, and the lower four panels show the percent difference for each of the satellite products. Note the TROPOMI data record began in May 2018. The scatter in the differences may be attributed to solar and viewing zenith angles, instrument degradation, presence of clouds, and other factors.

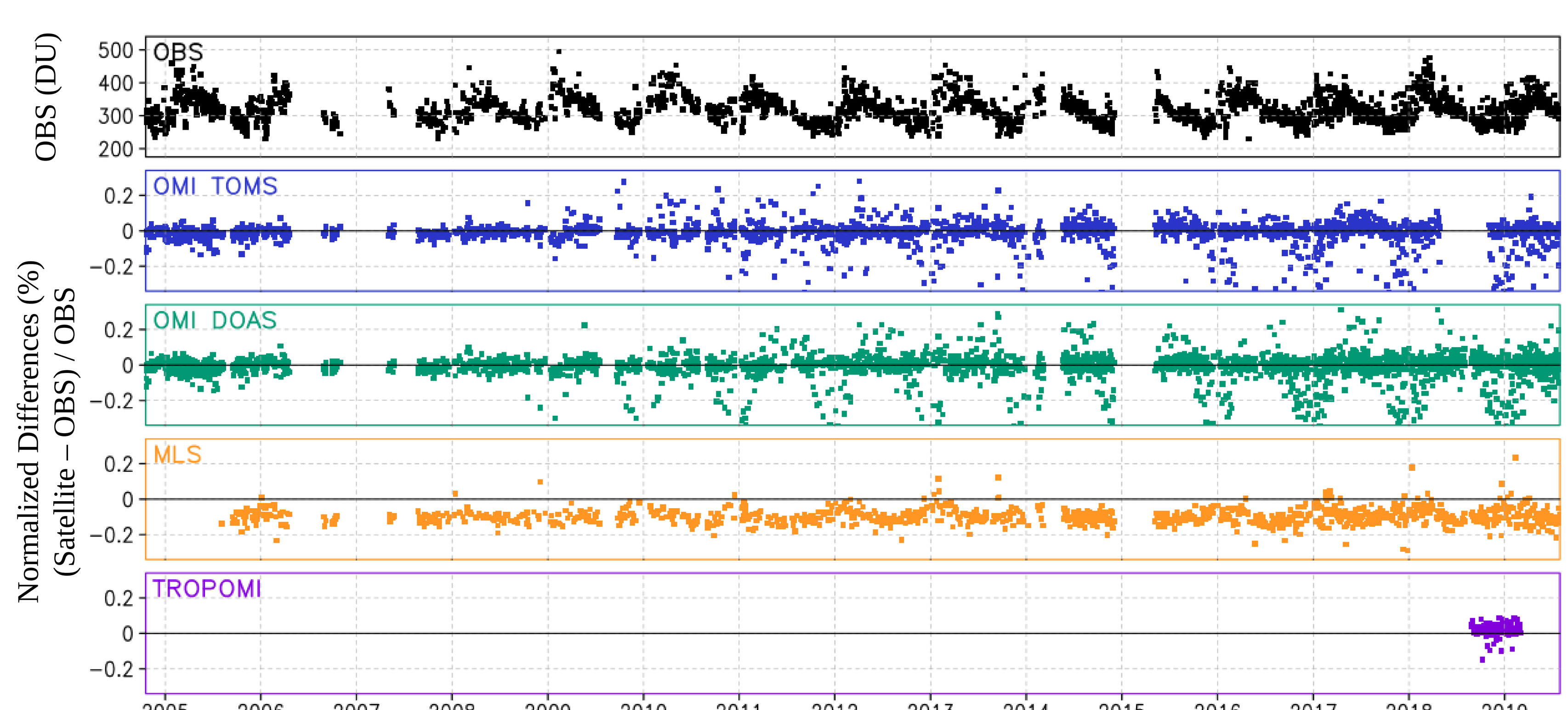

Figure 4. Time series of total ozone at Haute Provence, plus the percent difference for each co-located satellite measurement.

By utilizing the Level-2 subsetting service and the API, end users can automate data collection and significantly reduce downloaded data volume. The Level-2 subsets result in faster downloads due to a decrease in required data bandwidth, less local disk space required for storage and faster loading into applications for data analysis. Specifics about the resulting data volume and the percentages saved are in Table 3 below.

	OMI-TOMS			OMI-DOAS			TROPOMI			MLS		
Station	# Obs	Total Size	File Savings	# Obs	Total Size	File Savings	# Obs	Total Size	File Savings	# Obs	Total Size	File Savings
Utqiagvik (Barrow)	1232	282 MB	99.3%	1305	282 MB	98.5%	104	28 MB	99.9%	1301	295 MB	82%
Haute Provence	2464	565 MB	99.3%	2472	590 MB	98.4%	283	56 MB	99.9%	2730	303 MB	82%
South Pole	2304	518 MB	99.3%	2400	507 MB	98.5%	143	47 MB	99.9%			

Table 3. A summary of the file size savings for data granule subsets compared to the original granule size.

Summary

GES DISC Data Services simplify the workflow for Level-2 data acquisition

- ✓ Use **Search API** to find satellite data granules within the domain of interest
- ✓ Leverage the **Subset API** to generate the URLs for data variable extraction in preferred spatiotemporal regions
- ✓ Execute the **Level 2 Subsetting** calls embedded in the URLs that subset the selected granules
- ✓ **Download** only the data required for study
- ✓ Users **save time, effort, and resources**

Acknowledgements

The authors thank the GES DISC Level-2 Scientists and Engineers. Icon images obtained from Wikimedia Commons. Ground ozone data obtained from NOAA Earth System Research Laboratory Global Monitoring Division at esrl.noaa.gov. Satellite data products are available at disc.gsfc.nasa.gov.