Questions & Answers Part 2

Please type your questions in the Question Box. We will try our best to get to all your questions. If we don't, feel free to email Carl Malings (carl.a.malings@nasa.gov).

Question 1: Would stricter air quality standards (NAAQS), administered by the EPA, show more widespread cases of underserved communities across the United States, and would that have a more profound effect on government action to resolve these cases of EJ?
Answer 1: For those studying EJ issues, the exact air quality standards are typically less relevant than the differences in exposure observed for different population groups. However, as you suggest, more stringent air quality standards might have an impact on the kind of actions which would be taken to address pollution exposure, including disparities in exposure.

Question 2: Which satellites are in polar orbit?
Answer 2: There are many satellites in polar orbits; in the training I've mentioned a few which are most relevant to air quality. These include the Aqua and Terra satellites, which host the MODIS instrument, the Aura satellite, with the OMI instrument, the SNPP, NOAA-20, and NOAA-21 satellites with the VIIRS instrument, and the S5P satellite with the TROPOMI instrument.

Question 3: Does NASA have any open source models that can be loaded into Python for real time analysis?
Answer 3: If you are referring to atmospheric chemistry models, I'm not aware of any which can be run in Python in real time; these models are extremely large and complex, often requiring large computer clusters to run. For example, the NASA Global Modeling and Assimilation Office runs several operational global atmospheric models in the GEOS family. These models are openly available, but still require a lot of expertise and resources to effectively run. If you are referring to data analysis models, including machine learning models, I'd recommend checking out ARSET's Fundamentals of Machine Learning for Earth Science course. We'll also be giving an example of using Python to analyze datasets for EJ investigations in Part 3 of this training.

Question 4: Could you please explain the hyperspectral and multispectral differences again in equations?
Answer 4: The exact radiative transfer equations describing hyperspectral and multispectral imagery and retrievals are complex, and beyond the scope of this training. Conceptually, hyperspectral data, by providing more information about the intensity of light at different frequencies, allows us to have more "degrees of freedom" in our data to exploit when designing mathematical algorithms to extract different kinds of information from the satellite data. As an analogy, detecting different gasses in the atmosphere is like looking at a picture of a fingerprint and trying to figure out whose finger made that fingerprint; if you have a shaper image of the fingerprint (a hyperspectral image in this analogy), you will have an easier time to figure out whose finger made the print than if you only had a blurry image (a multispectral image in this analogy). I'd recommend reviewing the ARSET Fundamentals of Remote Sensing for more information about spectral resolution. A university course on optics, remote sensing, and radiative transfer theory would probably be required for the equations underlying these concepts.

Question 5: What would be a good resource to learn more about row anomalies and changes in older satellites?
Answer 5: Generally, the webpages of satellite missions & instruments will have sections discussing the performance of the instrument and any degradation which has been observed; for example. The instrument teams also publish scientific papers reporting on the performance of their instruments and any changes over time. For the OMI row anomaly specifically, Part 1 of our training on TROPOMI covers OMI and discusses the row anomaly.

Question 6: What would be a good resource to learn more about impact of smoke on satellite data (like what extent of smoke blocks the instrument)?
Answer 6: You can refer back to our recent ARSET training on MODIS and VIIRS, which goes into more detail about satellite aerosol retrievals and their limitations. Our training on wildfires, specifically Part 3, addresses this question as well, but focuses on smoke impacts on the detection of active fires. In general, areas of cloud cover and thick smoke will be given lower quality flags and have larger uncertainties associated with them, or be filtered out entirely.

Question 7: Will MAIA be able to provide vertical/3d air column info?
Answer 7: Some information about the vertical profiles of aerosols may be available from MAIA, but it is not a main focus of the mission. I'd suggest visiting the MAIA webpage for more information. If you are interested in aerosol vertical profiles, the CALIPSO mission might be of interest to you. The CALIPSO science mission has unfortunately officially ended as of August 1, 2023, but the historical record is available.

Question 8: Will any new or upcoming missions provide good methane emissions info that could help identify leaks and releases?
Answer 8: The EMIT mission is NASA's main mission looking at methane, including the detections of plumes indicating potential leaks. TROPOMI also has a methane data product, but this is better suited for regional studies of methane rather than searching for specific leaks.

Question 9: Are all these sensors passive sensors? There is no nighttime data collection yet?
Answer 9: Yes, all the sensors I've presented about today are passive sensors. There are some active sensors providing information on atmospheric composition, for example the CALIPSO mission, but they typically have very narrow swath widths, making them less useful for near-real time air quality applications. That mission also ended recently, although the historical record could be useful if the swath intersected your event of interest. Nighttime data collection on atmospheric composition is also not common; there has been some very interesting recent work attempting to get AOD information using moonlight instead of sunlight, but this is still in a preliminary phase.

Question 10: Could MISR be useful for measuring particles, just like MODIS does?
Answer 10: Yes, MISR is useful for gathering information about atmospheric particles. Like the upcoming MAIA mission, MISR collects information at multiple angles, which can be used to infer properties of aerosols, which can be useful for information about aerosol type, size, and altitude. While MISR also measures AOD similar to MODIS, it has a smaller swath width, so does not provide coverage every 1-2 days. Putting data from both instruments together can give us a better understanding of aerosols overall.

Question 11: Can datasets be overlayed in Giovanni like you did in Worldview?

Answer 11: You can select multiple datasets in Giovanni, and some types of plots (e.g., difference plots and correlation plots) allow you to compare two datasets with each other. However, you typically can't "overlay" images onto the same plot, like was possible in Worldview.

Question 12: When will the new North America satellite data be available?
Answer 12: If you are referring to TEMPO, it launched in April of 2023, and data should be released early next year. In fact, the TEMPO team just released some of its preliminary "first light" imagery.

Question 13: How can one improve Air Pollution retrieval techniques by artificial intelligence and machine learning algorithms? Is there any Case Study?
Answer 13: Most satellite retrieval algorithms make use of radiative transfer models to perform their retrievals. There has been research into replacing either the entire transfer retrieval algorithm or its most computationally intensive parts with machine learning surrogate models. Here is an example of a recent paper, testing this approach for OMI SO_2. Here is a review paper giving a broader view of the applications of machine learning to Earth Science. I'd also recommend checking out ARSET's Fundamentals of Machine Learning for Earth Science course.

Question 14: Satellite measurements generate some uncertainty, especially in areas of complex orography. This leads to few valid satellite measurements or significant quality control. How could it be handled or what do you suggest to handle this data in a $PM_{2.5}$ particulate material investigation using 550 nm AOD?
Answer 14: In converting between satellite 550 nm AOD and surface $PM_{2.5}$, there are many sources of uncertainty, including uncertainties in the AOD itself due to instrument noise and retrieval algorithm uncertainties, uncertainty in the relationships between column AOD and surface $PM_{2.5}$ concentrations, and uncertainties in how well the conditions when there are valid satellite data (during satellite overpass times during the day under cloud-free conditions) represent the conditions at other times (at night, when the satellite is not overhead, when it is cloudy, etc.). Averaging satellite data for longer periods of time is generally recommended to provide more cloud-free observations and to reduce the impacts of random noise; however, this will not reduce systematic sources of uncertainty (e.g., the fact that the satellite is passing overhead at the same local time every day). For that, other

sources of information (models, ground-based measurements) need to be considered as well. This paper describes the creation of a "Level 4" $PM_{2.5}$ data product using AOD and other types of data, gives an overview of how uncertainties were quantified by that approach, to give you a general idea of how this can be handled.

Question 15: SEDAC has $PM_{2.5}$ daily and annual at a 1 km resolution, but this set ends in 2016. Is there a time frame for updates to that set or any suggestions on algorithms available to update these models with available data?

Answer 15: In general, these "Level 4" data products require a lot of analysis to produce, and so lag significantly behind real time. The Atmospheric Composition Analysis Group at Washington University of St. Louis produces a dataset covering through 2021, which may be of interest to you. Although these research groups will typically update their datasets periodically, the frequency of doing so depends on the availability of research funding.

Question 16: I loved this part of the training! In the beginning, you had discussed the on-ground monitors and the discrepancies between the governmental approved and calibrated monitors and the cheap monitors. How could the AOD analysis show which monitors are more accurate?

Answer 16: This is a great question, and an open area of research. Generally, a regulatory monitor (such as one operated by a government agency) has extensive data quality control procedures to make sure that it is providing accurate data. However, because of the expenses associated with these quality control procedures, typically only a few monitors are deployed. The question is, therefore, whether the government monitors are representing the air quality situation throughout an entire region, or whether there are spatial patterns and trends which they are not capturing. Best practices for using low-cost sensors to assess this is to first co-locate the sensors with available government monitors, to establish what the data quality of the low-cost sensors are before using them to study spatial patterns and trends. However, there are many challenges to this, particularly for regions and nations which lack regulatory monitors, or only have one or two in a few major cities. This is where satellite data can play a role by establishing a "linkage" connecting the few government monitors with low-cost sensors deployed in other areas. However, there is not an established methodology for this, and it is an area of ongoing research.

My advice to address this question would be to adopt a "weight of evidence" approach, inter-comparing all available data sources and looking for situations where they all agree on common trends and patterns. For example, if you are detecting the same patterns and trends using low-cost sensors and satellite AOD, which are two independent data sources, you have more confidence to believe that these observed patterns and trends are actually occurring.

Question 17: Does NASA have any of these datasets available as OGC map services or other types of REST services that could be easily and directly loaded into web maps or tools like ArcGIS Pro without having to download and process the data first?
Answer 17: Although we aren't aware of specific efforts to create OGC maps with NASA data, the NASA Transform to Open Science initiative aims to make NASA data more accessible and interoperable using other analysis tools. The NASA VEDA project aims to facilitate NASA data analysis and plotting in GIS systems, and NASA Earthdata hosts a collection of web-based GIS applications, including one of the Level 4 $PM_{2.5}$ datasets we discussed. Various NASA datasets are also available on Google Earth Engine. There is also the Earthdata API to automatically query NASA databases for specific data (we showed an example of how to do this manually during Part 2 of the training; using the API can automate that process).

Question 18: How can I convert the OMI tropospheric NO_2 unit in molecules/cm2 for ppbv to compare with NO_2 surface measured data?
Answer 18: The relationship between the NO_2 tropospheric column and surface concentrations depends on the vertical distribution of NO_2 within the column. In general, another information source like a model or surface-based measurements of NO_2 are needed to establish a relationship between these two quantities. The ARSET training on NASA Air Quality-Focused Remote Sensing for EPA Applications includes some examples of techniques that have been used to relate NO_2 column and surface observations. Here is an example applied to TROPOMI data. You can also use a pre-computed Level 4 dataset, such as this one for the US or the global dataset developed by the GWU group discussed in this training.

Question 19: How accurate is the air quality data collected from satellites, especially in areas with varying environmental conditions? What are the main sources of uncertainty in satellite-based air quality measurements, and how are they addressed?
Answer 19: There are several sources of uncertainty. No instrument is perfect, and satellite measurements of radiance are subject to some uncertainties. Some additional uncertainty is introduced when converting what the satellite is seeing (the atmospheric column) to air

quality on the ground level. Satellites also have limited coverage due to factors such as clouds, night, and orbital return periods. Averaging over long periods of time will help with some sources of uncertainty, but not others (e.g., systematic biases due to the satellite only observing at specific times of day). Bringing in other sources of information, especially surface-based observations (where available) and models, can help to characterize this uncertainty. This paper describes the creation of a "Level 4" $PM_{2.5}$ data product using satellite AOD and other types of data, gives an overview of how uncertainties were quantified by that approach, to give you a general idea of how this might be handled.

Question 20: How do satellite-derived air quality measurements complement or differ from ground-based monitoring stations? Are there challenges in reconciling data from satellite observations with ground-level measurements?

Answer 20: Absolutely. Satellite and ground-based data can complement each other well. Ground-based measurements are the best estimate of local air quality, but do not exist everywhere. Satellite measurements help bring wide-area or even global coverage, but are limited (observing the entire atmosphere and not just near the surface, observing only during specific times of day, etc.). Bringing these data together and using the strengths of one to offset the weaknesses of the other is a good approach to developing a better understanding of the air quality station overall, but requires collecting a lot of data and making some assumptions about how to best combine them together.

Question 21: Would you recommend combining the AOD data from Terra and Aqua for better $PM_{2.5}$ resolution, apart from contrasting low-cost and ground measurements?

Answer 21: Combining AOD data from the Terra and Aqua satellites is already done with the MAIAC data product. You need to be careful to consider that the satellites are passing overhead at different times, and so are not looking at precisely the same thing, but the data combination can still be done. Several Level 4 $PM_{2.5}$ data products use data from both Aqua and Terra (together with other information sources) as inputs; for example, the MAIAC data product is one input to the methodology described here.

Question 22: Can LiDAR be used to help fill the gap between 2.2 KM and Ground monitoring?

Answer 22: Yes. LiDAR gives a lot of information on vertical measurements where this is available, but they are only available in a few specific places. A model can help bridge these observations. Here is an example where CALIPSO LiDAR data was used to estimate surface $PM_{2.5}$ concentrations.